

STIC-Biotech/ChemLib

170010

From: Walicka, Malgorzata  
Sent: Saturday, October 29, 2005 10:56 AM  
To: STIC-Biotech/ChemLib

Please search SEQ ID NO:1 and 16 in 09/774,954. Please use also pending data.

Thank you.

Malgorzata A. Walicka, Ph.D.  
Patent Examiner  
Art Unit 1652, Recombinant Enzymes  
USPTO, Remsen Building, Room 2C76  
400 Dulany St.  
Alexandria, VA 22313  
Mail Room 2C70  
Tel. (571) 272-0944, fax (571) 273-0944

\*\*\*\*\*

Searcher: \_\_\_\_\_  
Searcher Phone: \_\_\_\_\_  
Date Searcher Picked up: 11/11/05  
Date completed: 11/19/05  
Searcher Prep Time: \_\_\_\_\_  
Online Time: \_\_\_\_\_

\*\*\*\*\*

Type of Search  
NA# 2 AA#: \_\_\_\_\_  
S/L: \_\_\_\_\_ Oligomer: \_\_\_\_\_  
Encode/Transl: \_\_\_\_\_  
Structure #: \_\_\_\_\_ Text: \_\_\_\_\_  
Inventor: \_\_\_\_\_ Litigation: \_\_\_\_\_

\*\*\*\*\*

Vendors and cost where applicable  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
QUESTEL/ORBIT: \_\_\_\_\_  
LEXIS/NEXIS: \_\_\_\_\_  
SEQUENCE SYSTEM: 04  
WWW/Internet: \_\_\_\_\_  
Other (Specify): \_\_\_\_\_

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2005, 23:53:44 ; Search time 6830.38 Seconds  
(without alignments)  
10740.434 Million cell updates/sec

Title: US-09-774-954-1  
Perfect score: 1514  
Sequence: 1 ATGCCCGCGGCTCTCGGA.....TTTGAATTCGAATTC 1514

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_hgt.\*

3: gb\_in.\*

4: gb\_on.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pl.\*

9: gb\_pt.\*

10: gb\_ro.\*

11: gb\_sts.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vi.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1514	100.0	1514	6	ARI63451 Sequence
2	1514	100.0	1514	6	BD103188 O-fucosyl
3	1512.4	99.9	5249	9	AF375884 Homo sapi
4	1512.4	99.9	11284	6	BD103190 O-fucosyl
5	1510.8	99.8	5189	9	D80002 Homo sapien
6	1509.8	99.7	5266	6	AX780153
7	1507.6	99.6	4560	9	HSM803330
8	1447.4	95.6	5009	6	ARI63453
9	1447.4	95.6	5009	6	BD103191 O-fucosyl
10	1163.4	76.8	1300	6	ARI63452 Sequence
11	1163.4	76.8	1300	6	BD103189 O-fucosyl
12	1100	72.7	1100	6	ARI63457 Sequence
13	1100	72.7	1100	6	BD103196 O-fucosyl
14	1096.4	72.4	1167	6	CQ727777 Sequence
15	1086.8	71.8	1167	9	AJ781500 Pan trogl
16	916.2	60.5	1176	4	AY344580 Bos tauru
17	908.2	60.0	1176	4	AJ781503
18	901.6	59.6	1134	4	SSC567917 Sus scrofa
19	889.6	58.8	1188	10	AJ781499 Rattus no

20	889.6	58.8	1302	10	AF375885
21	889.6	58.8	1529	10	BC046295
22	752.8	49.7	1056	4	AY344581 Bos tauru
23	652.2	43.1	1732	5	GGAS35754
24	652.2	43.1	2546	5	AJ720352 Gallus ga
25	608.6	40.2	145414	9	HS4392M18
26	586.4	38.7	1502	5	BC082519
27	586.4	38.7	1538	5	AJ781498
28	563.6	37.2	1155	5	AJ606070 Fugu rubr
29	552.8	36.5	1505	5	XLA514425
30	539	35.6	1164	5	AJ781504
31	538.4	35.6	1549	5	DRE579536
32	471.4	31.1	1524	9	BC000582
33	445.4	29.4	920	5	AJ719585
34	201.6	13.3	1841	3	AK112708
35	193.4	12.8	2393	9	AK124093
36	188.8	12.5	240336	2	AC094844
37	186.6	12.3	240510	2	AC134371
38	184.6	12.2	200329	10	AC078911
39	184.6	12.2	208614	10	AL807380
40	184	12.2	1146	3	AJ781502
41	172.8	11.4	1209	3	AJ831490
42	168	11.1	1295	3	AY118651
43	168	11.1	1320	6	ARS09393
44	166.8	11.0	1209	3	AJ781501
45	166.4	11.0	1209	6	CQ590403

ALIGNMENTS

RESULT 1

ARI63451

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

ORIGIN

Sequence 2 from patent US 6270987.  
ARI63451  
ARI63451.1 GI:16234056

Unknown.

Unclassified.

1 (bases 1 to 1514)

Wang, Y. and Spellman, M.W.

O-fucosyltransferase

Patent: US 6270987-A 2 07-AUG-2001;

Location/Qualifiers

1..1514

/organism="unknown"

/mol\_type="unassigned DNA"

Query Match 100.0%; Score 1514; DB 6; Length 1514;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1514; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGCCCGCGGCTCTCTGGGACCCGGCGGTTACTGTCTACTGCTCCCTGATGGGGCGC 60  
Db 1 ATGCCCGCGGCTCTCTGGGACCCGGCGGTTACTGTCTACTGCTCCCTGATGGGGCGC 60

Qy 61 TTTGGGAACCAAGGCCGATCACTTCTTGGGCTCTCTGGCAATTTGCAAGCTGCTAAACCGT 120

Db 61 TTTGGGAACCAAGGCCGATCACTTCTTGGGCTCTCTGGCAATTTGCAAGCTGCTAAACCGT 120

Qy 121 ACCTTGGCTGCTCCCTCTGGATTGAGTACACAGCTCACAGCTCCCTTTCCACCAACCTC 180

Db 121 ACCTTGGCTGCTCCCTCTGGATTGAGTACACAGCTCACAGCTCCCTTTCCACCAACCTC 180

Qy 181 CATGTGCTTACCAAGAGTACTTCAAGCTGAGGCCCTCCAGGCTTACCATCGGCTCATC 240

Db 181 CATGTGCTTACCAAGAGTACTTCAAGCTGAGGCCCTCCAGGCTTACCATCGGCTCATC 240

Qy 241 AGCTTGGAGGATTTTCATGGAGAACTGGCACCACCACTGGCCCTTGAGAGCGGTG 300

Db 241 AGCTTGGAGGATTTTCATGGAGAACTGGCACCACCACTGGCCCTTGAGAGCGGTG 300

Db 241 AGCTTGGAGGATTTTCATGGAGAGCTGGCACCACCCACCTGGCCCCCTTGAGAGCGGGTG 300  
Qy 301 GCATATGCTCTTTGAGGTGGCAGCCAGCGAAGCCAGATGAAGAAGCTGGCCCCATGAAG 360  
Db 301 GCATATGCTCTTTGAGGTGGCAGCCAGCGAAGCCAGATGAAGAAGCTGGCCCCATGAAG 360  
Qy 361 GAAGGAACCCCTTTGGCCCATCTTGGGATCAGTTCCTGAGTTCATGAGTTCACAAAGTCGGAG 420  
Db 361 GAAGGAACCCCTTTGGCCCATCTTGGGATCAGTTCCTGAGTTCATGAGTTCACAAAGTCGGAG 420  
Qy 421 CTTTTCACAGCATTTCTTCTGAGTTCCTTCTGAGTTCCTTCTGAGTTCCTTCTGAGTTCCTTCT 480  
Db 421 CTTTTCACAGCATTTCTTCTGAGTTCCTTCTGAGTTCCTTCTGAGTTCCTTCTGAGTTCCTTCT 480  
Qy 481 CCAAGGAAGATCCGCTGCTTCTGAGTTCCTTCTGAGTTCCTTCTGAGTTCCTTCTGAGTTCCTTCT 540  
Db 481 CCAAGGAAGATCCGCTGCTTCTGAGTTCCTTCTGAGTTCCTTCTGAGTTCCTTCTGAGTTCCTTCT 540  
Qy 541 GAACACAGGCCACTACAGAAAGTACATGATGATGATGATGATGATGATGATGATGATGATGATGATG 600  
Db 541 GAACACAGGCCACTACAGAAAGTACATGATGATGATGATGATGATGATGATGATGATGATGATGATG 600  
Qy 601 GCCAGATTCATGCCCCACCTTGTGCGGCCCTATGCGGCCATTCATCTGCGCATTCGGCTCT 660  
Db 601 GCCAGATTCATGCCCCACCTTGTGCGGCCCTATGCGGCCATTCATCTGCGCATTCGGCTCT 660  
Qy 661 GACTGGAAGAGCGCTGTGCGCATCTGCGGCCCTATGCGGCCATTCATCTGCGCATTCGGCTCT 720  
Db 661 GACTGGAAGAGCGCTGTGCGCATCTGCGGCCCTATGCGGCCATTCATCTGCGCATTCGGCTCT 720  
Qy 721 TCTCGCAGTGTGTGGCTACAGCCGACGACGAGCGGCCCTTCTGCGCATTCGGCTCT 780  
Db 721 TCTCGCAGTGTGTGGCTACAGCCGACGACGAGCGGCCCTTCTGCGCATTCGGCTCT 780  
Qy 781 CTGCGTACCTGAGGAGATCCAGAGGCTGTGAGGCTCTGCGGCCATTCGGCTCT 840  
Db 781 CTGCGTACCTGAGGAGATCCAGAGGCTGTGAGGCTCTGCGGCCATTCGGCTCT 840  
Qy 841 CAGTGGCTACGTTGCTTCTGAGTTCCTGAGTTCCTGAGTTCCTGAGTTCCTGAGTTCCTGAGTTCCT 900  
Db 841 CAGTGGCTACGTTGCTTCTGAGTTCCTGAGTTCCTGAGTTCCTGAGTTCCTGAGTTCCTGAGTTCCT 900  
Qy 901 AAGGGAAGGTGAAGGTGGTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCT 960  
Db 901 AAGGGAAGGTGAAGGTGGTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCT 960  
Qy 961 CTGCGCAAGCGGACCACTTATTTGGCAACTGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1020  
Db 961 CTGCGCAAGCGGACCACTTATTTGGCAACTGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1020  
Qy 1021 CGGAGCGGACCTCCAGGGGAGGCGCTCTTCTTCTTCTGCGCATGAGGCGGCGGCGGCGGCGGCGG 1080  
Db 1021 CGGAGCGGACCTCCAGGGGAGGCGCTCTTCTTCTTCTGCGCATGAGGCGGCGGCGGCGGCGGCGG 1080  
Qy 1081 CTGCGGAGCGAGTTCGATTCGCGGAGCACAGACCTCTGATCTGCGGAGGAGGAGGAGGAGGAGGAG 1140  
Db 1081 CTGCGGAGCGAGTTCGATTCGCGGAGCACAGACCTCTGATCTGCGGAGGAGGAGGAGGAGGAGGAG 1140  
Qy 1141 AGTCTGAGCTGCTCTTCCAGCGGAGGCGGCTTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1200  
Db 1141 AGTCTGAGCTGCTCTTCCAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1200  
Qy 1201 CTTCTTCTCAGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1260  
Db 1201 CTTCTTCTCAGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1260  
Qy 1261 TCCATATCCAGGCGATAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1320  
Db 1261 TCCATATCCAGGCGATAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1320  
Qy 1321 GTGCTTCT 1380  
Db 1321 GTGCTTCT 1380

Qy 1381 GGTCCACTCTGCTCTGAGCAGCTGGGATGCTGAACCTTTCAGAGAGATTTTATAGA 1440  
Db 1381 GGTCCACTCTGCTCTGAGCAGCTGGGATGCTGAACCTTTCAGAGAGATTTTATAGA 1440  
Qy 1441 GAGATTTCTATAAATTTTGATACAAAGTCATGATCTCTAGAACTCTCTGTGTTTTTGA 1500  
Db 1441 GAGATTTCTATAAATTTTGATACAAAGTCATGATCTCTAGAACTCTCTGTGTTTTTGA 1500  
Qy 1501 AAATCATTTGAATTC 1514  
Db 1501 AAATCATTTGAATTC 1514  
  
RESULT 2  
BD103188  
LOCUS BD103188 1514 bp DNA linear PAT 27-AUG-2002  
DEFINITION O-fucosyltransferase.  
ACCESSION BD103188.1 GI:22648762  
VERSION BD103188.1  
KEYWORDS JF001527389-A/1.  
SOURCE unclassified  
ORGANISM unclassified  
REFERENCE 1 (bases 1 to 1514)  
AUTHORS Wang, Y. and Spellman, M. W.  
TITLE O-fucosyltransferase  
JOURNAL Patent: JP 2001527389-A 1 25-DEC-2001;  
GENENTECH INC  
COMMENT OS Unidentified  
PN JP 2001527389-A/1  
PD 25-DEC-2001  
PF 17-DEC-1997 JP 1998532877  
PR 31-JAN-1997 US 08/792498, 26-NOV-1997 US 08/978741 PI  
YANG WANG, MICHAEL W SPELLMAN  
PC C12N15/54, C12N9/10, C07K16/40  
CC Strandedness: Single;  
CC Topology: Linear;  
CC O-fucosyltransferase  
FH Key Location/Qualifiers  
FT source 1..1514  
FT location/Qualifiers  
FT 1..1514 /organism='Unidentified'.  
  
FEATURES  
source  
1..1514  
/organism='unidentified'  
/mol\_type='genomic DNA'  
/db\_xref='taxon:32644'  
  
ORIGIN  
Query Match 100.0%; Score 1514; DB 6; Length 1514;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1514; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 ATCCCGCGGGCTCTGGGACCCCGCGGTTACCTGCTCTACTGCCCCCTGCATGGGGCGC 60  
Db 1 ATCCCGCGGGCTCTGGGACCCCGCGGTTACCTGCTCTACTGCCCCCTGCATGGGGCGC 60  
Qy 61 TTTGGGAACCAAGCCGATCACTCTTGGGCTCTCTGGCAATTGCAAAAGCTGCTAAACCGT 120  
Db 61 TTTGGGAACCAAGCCGATCACTCTTGGGCTCTCTGGCAATTGCAAAAGCTGCTAAACCGT 120  
Qy 121 ACCTTGGCTGCTCCCTCTTGGATTGAGTACCAAGCATCAAAAGCTTCTTTCAGCAACCTC 180  
Db 121 ACCTTGGCTGCTCCCTCTTGGATTGAGTACCAAGCATCAAAAGCTTCTTTCAGCAACCTC 180  
Qy 181 CATGCTCTTACCAAGATCTTCAAGCTGGAGCCCTCCAGGCTTACCTGCGGTCATC 240  
Db 181 CATGCTCTTACCAAGATCTTCAAGCTGGAGCCCTCCAGGCTTACCTGCGGTCATC 240  
Qy 241 AGTTGGAGGATTTTCAGAGAGCTGGACACCCACCCCTGSCCCCTGAGAGCGGGTG 300  
Db 241 AGTTGGAGGATTTTCAGAGAGCTGGACACCCACCCCTGSCCCCTGAGAGCGGGTG 300



```
QY 301 GCATCTGCTTTGAGTGGCAGCCAGCAAGCCAGATAGAAAGACGTGCCCATGAAG 360
DB |||||
QY 301 GCATCTGCTTTGAGTGGCAGCCAGCAAGCCAGATAGAAAGACGTGCCCATGAAG 360
DB |||||
QY 361 GAAGAAACCCCTTTGGCCCATCTTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
DB |||||
QY 361 GAAGAAACCCCTTTGGCCCATCTTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
DB |||||
QY 421 CTTTTTACAGGCAATTTCTTCTAGTCTCTTACAGAAACATGAGGACGAGATTTTCT 480
DB |||||
QY 421 CTTTTTACAGGCAATTTCTTCTAGTCTCTTACAGAAACATGAGGACGAGATTTTCT 480
DB |||||
QY 481 CCAAGGAACATCCGGTCTTCCCTGCCAGGAGCCAGCCAGTTCCTCCGTCCTAGAA 540
DB |||||
QY 481 CCAAGGAACATCCGGTCTTCCCTGCCAGGAGCCAGCCAGTTCCTCCGTCCTAGAA 540
DB |||||
QY 541 GAACACAGGCCCATCAGAAAGTACATGTTATGTTGAGTGAAGACGAGGAG 600
DB |||||
QY 541 GAACACAGGCCCATCAGAAAGTACATGTTATGTTGAGTGAAGACGAGGAG 600
DB |||||
QY 601 GCCCAGATTCATGCCACCTTGTCCGCGCCCTATGTGGGCATTCATCTGGCATTTGGCTCT 660
DB |||||
QY 601 GCCCAGATTCATGCCACCTTGTCCGCGCCCTATGTGGGCATTCATCTGGCATTTGGCTCT 660
DB |||||
QY 661 GACTCGAAGAACCGCTGTGCCATGCTGAAGACGGGACTGCAAGGCTCGCACTTCATGACC 720
DB |||||
QY 661 GACTCGAAGAACCGCTGTGCCATGCTGAAGACGGGACTGCAAGGCTCGCACTTCATGACC 720
DB |||||
QY 721 TCTCCGAGTGTGGGTACAGCCGACAGCAGCGGCCCTCCTCAAGTACATATGTC 780
DB |||||
QY 721 TCTCCGAGTGTGGGTACAGCCGACAGCAGCGGCCCTCCTCAAGTACATATGTC 780
DB |||||
QY 781 CTGCTGACCTGAAGGAGATCAGAGGCTGTGAAGCTCTGGTCTGGTCTGGTCTGGTCTGG 840
DB |||||
QY 781 CTGCTGACCTGAAGGAGATCAGAGGCTGTGAAGCTCTGGTCTGGTCTGGTCTGGTCTGG 840
DB |||||
QY 841 CAGTCGGTCTAGCTTACTGATTCAGAGAGTATGTGCTGCTGAGCTCCAAAGCTCTTC 900
DB |||||
QY 841 CAGTCGGTCTAGCTTACTGATTCAGAGAGTATGTGCTGCTGAGCTCCAAAGCTCTTC 900
DB |||||
QY 901 AAAGGAAGGTGAAGGTGTGAGCTGAAGCTGTAGGTGGCCAGGTGCACTGTACATC 960
DB |||||
QY 901 AAAGGAAGGTGAAGGTGTGAGCTGAAGCTGTAGGTGGCCAGGTGCACTGTACATC 960
DB |||||
QY 961 CTGCGCAGCGGACCACTTTATTTGGCACTGTGTCTCTCTTCACTGCCTTTGTGAG 1020
DB |||||
QY 961 CTGCGCAGCGGACCACTTTATTTGGCACTGTGTCTCTCTTCACTGCCTTTGTGAG 1020
DB |||||
QY 1021 CGGAGCGGACCTCCAGGGAGGCGCTCTTCTTCTTCCGATGGAAGAGGCGCCCTAAG 1080
DB |||||
QY 1021 CGGAGCGGACCTCCAGGGAGGCGCTCTTCTTCTTCCGATGGAAGAGGCGCCCTAAG 1080
DB |||||
QY 1081 CTGCGGACGAGTCTGTATTTGGCCGAGACCAAGCCCTCTGTATCTTGAGGAGACAG 1140
DB |||||
QY 1081 CTGCGGACGAGTCTGTATTTGGCCGAGACCAAGCCCTCTGTATCTTGAGGAGACAG 1140
DB |||||
QY 1141 AGTCTGAGCTGCTCTTCCAGCCAGCTTGGCAGCAGAGTGTCTCCGGGATTCGAAC 1200
DB |||||
QY 1141 AGTCTGAGCTGCTCTTCCAGCCAGCTTGGCAGCAGAGTGTCTCCGGGATTCGAAC 1200
DB |||||
QY 1201 CCTCTTCTCAGCTGCAAGATGGAAGAGAGTGCAGGAGCCCTCAAGGAGGAGAGCG 1260
DB |||||
QY 1201 CCTCTTCTCAGCTGCAAGATGGAAGAGAGTGCAGGAGCCCTCAAGGAGGAGAGCG 1260
DB |||||
QY 1261 TCCATATCCAGGGATAGGACTTGAGTCTTCTAGGAGCAGGAGCATCTCCCATCGCAC 1320
DB |||||
QY 1261 TCCATATCCAGGGATAGGACTTGAGTCTTCTAGGAGCAGGAGCATCTCCCATCGCAC 1320
DB |||||
QY 1321 GTGCTTCTGCTCTTCTGGAATTTCTCAGTGGCAAGCAGTCCAGCTCTGCTTCT 1380
DB |||||
QY 1321 GTGCTTCTGCTCTTCTGGAATTTCTCAGTGGCAAGCAGTCCAGCTCTGCTTCT 1380
DB |||||
QY 1381 GGTCCACTCTGCTCTGAGCAGCGCTGGGATGCTGAACCTCTTCAGAGAGATTTTTTTTATAGA 1440
DB |||||
```

```
DB 1381 GGTCCACTCTGCTCTGAGCAGCGCTGGGATGCTGAACCTCTTCAGAGAGATTTTTTTTATAGA 1440
QY 1441 GAGATTTCTATAATTTTGTATACAAGGTCAAGTCACTATCTAGAACTCTCTGTGTTTTTGA 1500
DB |||||
QY 1441 GAGATTTCTATAATTTTGTATACAAGGTCAAGTCACTATCTAGAACTCTCTGTGTTTTTGA 1500
DB |||||
QY 1501 AAATCATTGAATTC 1514
DB |||||
QY 1501 AAATCATTGAATTC 1514
DB |||||

RESULT 3
AF375884
LOCUS Homo sapiens protein o-fucosyltransferase (POFUT1) mRNA, complete
DEFINITION AF375884.1 GI:15825113
ACCESSION AF375884
VERSION AF375884.1
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 5249)
AUTHORS Wang, Y., Shao, L., Shi, S., Harris, R.J., Spellman, M.W., Stanley, P.
and Haltiwanger, R.S.
TITLE Modification of epidermal growth factor-like repeats with O-fucose.
Molecular cloning and expression of a novel GDP-fucose protein
O-fucosyltransferase
J. Biol. Chem. 276 (43), 40338-40345 (2001)
JOURNAL 21523965
MEDLINE 11524432
PUBMED 11524432
REFERENCE 2 (bases 1 to 5249)
AUTHORS Shi, S., Stanley, P., Wang, Y., Shao, L., Harris, R.J., Spellman, M.W.
and Haltiwanger, R.S.
TITLE Direct Submission
JOURNAL Submitted (02-MAY-2001) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave., New York, NY 10461, USA
FEATURES
source
1..5249
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="20"
/map="between PLAGL2 and KIF3B"
/tissue_type="heart"
1..5249
/gene="POFUT1"
50..1216
/gene="POFUT1"
/functions="catalyzes transfer of fucose from GDP-fucose to
serine or threonine on epidermal growth factor-like
repeats"
/notes="O-FucT-1; glycosyltransferase"
/codon_start=1
/product="protein o-fucosyltransferase"
/protein_id="AAL09576.1"
/db_xref="GI:15825114"
translating="MGAARWPLSVFLLLLPLPGMAGSDPAGYLLYCPQWGRF
GNQADHPLGSLAFKLLNRTLVAPWIEYOHKPPFTNLHVSYQYFKLEPQAVHRV
ISLDFWEKLAHPHPKRVAYCFEVAQORSPDKTCPKKEGPPGFPWQPVHSE
KSLFTGISASRYEQWSPKPEHPVLALPCAPQFPVLEHRPLQKYMVSDEN
VKTGEAQIHAHLVRYVGIHLRIGSDWKNACAMLDGTAGSHFNASPCQCVSRSTAA
PLTMTCLPKLKEIKRAVLKWLSDAQSVYVATDSYVPELQQLFKGKVKVSLKP
EVAQVDLYLLQADHFICNCSVSTAFVXRERDLQGRPSFPFGMDRPPKLDEF"

Query Match 99.9%; Score 1512.4; DB 9; Length 5249;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1513; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ATGCCCGCGGCTCTCTGGGACCGCGGCTTACCTGCTCTACTGCTCTACTGCCCTCGATGGGCGC 60
```

Db 119 ATGCCCGGGCTCTCTGGGACCGGCGGGTTACTGTCTACTGCGCCCTGATCGGGCGC 178  
Qy 61 TTTTGGGAACAGAGCCGATCACTTCTTTGGGCTCTCTGGCATTTTGCATAAGCTCTAAACCGT 120  
Db 179 TTTTGGGAACAGAGCCGATCACTTCTTTGGGCTCTCTGGCATTTTGCATAAGCTCTAAACCGT 238  
Qy 121 ACCTTGGCTGTCCCTCTCTTGGATTCAGTACCAAGCATCAAAAGCTCTCTTTTCAACAACTTC 180  
Db 239 ACCTTGGCTGTCCCTCTCTTGGATTCAGTACCAAGCATCAAAAGCTCTCTTTTCAACAACTTC 298  
Qy 181 CATGTGTCTTACAGAAAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGCTCATC 240  
Db 299 CATGTGTCTTACAGAAAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGCTCATC 358  
Qy 241 AGCTTGGAGGATTTTCAATGGAGCTGGCAACCACTTGGCCCTTGGAGAGCGGGT 300  
Db 359 AGCTTGGAGGATTTTCAATGGAGCTGGCAACCACTTGGCCCTTGGAGAGCGGGT 418  
Qy 301 GATATCTCTTTGAGTGGAGCCGAGCAAGCTCCAGATTAAGAAAGCTGCCCATGAG 360  
Db 419 GATATCTCTTTGAGTGGAGCCGAGCAAGCTCCAGATTAAGAAAGCTGCCCATGAG 478  
Qy 361 GAAGGAAACCCCTTTGGCCCATCTGGGATCAGTTTCAATGTAGTTTCAACAAAGTCGAG 420  
Db 479 GAAGGAAACCCCTTTGGCCCATCTGGGATCAGTTTCAATGTAGTTTCAACAAAGTCGAG 538  
Qy 421 CTTTTTACAGGCATTTCTCTTCAAGTCTTCTACAGAGCAAAATGGAGCCAGAGATTTTCT 480  
Db 539 CTTTTTACAGGCATTTCTCTTCAAGTCTTCTACAGAGCAAAATGGAGCCAGAGATTTTCT 598  
Qy 481 CCAAGGAACATCCGGTCTTGGCCCTGCGAGAGCCCGAGCTTCCCGCTCTAGAA 540  
Db 599 CCAAGGAACATCCGGTCTTGGCCCTGCGAGAGCCCGAGCTTCCCGCTCTAGAG 658  
Qy 541 GAACACAGCCCACTACAGAGTACATGTTTGGTTCAGACGAAATGTGAAGACGGAGAG 600  
Db 659 GAACACAGCCCACTACAGAGTACATGTTTGGTTCAGACGAAATGTGAAGACGGAGAG 718  
Qy 601 GCGCAGATGATGCGCACCTTGTCCGCTTATGTGGCATTCATCTGGCATTCGCTCT 660  
Db 719 GCGCAGATGATGCGCACCTTGTCCGCTTATGTGGCATTCATCTGGCATTCGCTCT 778  
Qy 661 GACTGGAAGAACGCTTGTCCATGCTGAAGACGGGACTGAGGCTCGCATTCATGCCC 720  
Db 779 GACTGGAAGAACGCTTGTCCATGCTGAAGACGGGACTGAGGCTCGCATTCATGCCC 838  
Qy 721 TCTCCGAGTGTGGGCTACAGCCGACACACAGCGGCCCTCTCAGATGATGTGC 780  
Db 839 TCTCCGAGTGTGGGCTACAGCCGACACACAGCGGCCCTCTCAGATGATGTGC 898  
Qy 781 CTGCTGACCTGAAGAGGATCCAGAGGCTGTGAAGCTCTGGGCTGAGGCTCGCTGGATGCC 840  
Db 899 CTGCTGACCTGAAGAGGATCCAGAGGCTGTGAAGCTCTGGGCTGAGGCTCGCTGGATGCC 958  
Qy 841 CAGTGGTCTAGTGTCTTACTGATTCGAGAGTTATGTGCTGAGCTCCCAACAGCTCTTC 900  
Db 959 CAGTGGTCTAGTGTCTTACTGATTCGAGAGTTATGTGCTGAGCTCCCAACAGCTCTTC 1018  
Qy 901 AAGAGGAGTGAAGTGTGAGCTGAGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGCTGATC 960  
Db 1019 AAGAGGAGTGAAGTGTGAGCTGAGCTGAGGCTGAGGCTGAGGCTGAGCTGATC 1078  
Qy 961 CTGCGCAAGCGGACCACTTTATGGCACTGTGTCTCTCTTCACTGCTGCTTGTGAAG 1020  
Db 1079 CTGCGCAAGCGGACCACTTTATGGCACTGTGTCTCTCTTCACTGCTGCTTGTGAAG 1138  
Qy 1021 CGGAGCGGAGCTTCCAGGGAGGCGCTTCTTTCTTGGCATGGAAGGCCCTCTTAA 1080  
Db 1139 CGGAGCGGAGCTTCCAGGGAGGCGCTTCTTTCTTGGCATGGAAGGCCCTCTTAA 1198  
Qy 1081 CTGCGGAGCGAGTCTTGAATCTGCGGAGCAACCACTCTGATCTCGAGGAGCAAG 1140

Db 1199 CTGCGGAGCGAGTTCTGATTTCTGGCCGGAGCACAGACCTCTGATCTCTGGAGGACCA 1258  
Qy 1141 AGTCTGAGCTGTCTCTTTCAGAGCCAGGCTTGGAGCTTGGAGTCTCTCGGATTTGCAAACT 1200  
Db 1259 AGTCTGAGCTGTCTCTTTCAGAGCCAGGCTTGGAGCTTGGAGTCTCTCGGATTTGCAAACT 1318  
Qy 1201 CTTCTTCTACCTGCTCCAAAGATGGAGAGTGGAGGACCCCTCAAGAGGGAGAGCG 1260  
Db 1319 CTTCTTCTACCTGCTCCAAAGATGGAGAGTGGAGGACCCCTCAAGAGGGAGAGCG 1378  
Qy 1261 TCCATATCCAGGCGCATAGGACTTTCAGGTTCTTAGGAGGAGGAGCATCTCCCATCGCAC 1320  
Db 1379 TCCATATCCAGGCGCATAGGACTTTCAGGTTCTTAGGAGGAGGAGCATCTCCCATCGCAC 1438  
Qy 1321 GTGCTTCTGCTCTTCTGGGAAATTTCTCACACTGGCAAGCAGTCCAGGCTCCGCTTCT 1380  
Db 1439 GTGCTTCTGCTCTTCTGGGAAATTTCTCACACTGGCAAGCAGTCCAGGCTCCGCTTCT 1498  
Qy 1381 GTGCCACTTCTCTCTGAGCAGCTGGGATGCTGAACTCTTTCAGAGAGATTTTATAGA 1440  
Db 1499 GTGCCACTTCTCTCTGAGCAGCTGGGATGCTGAACTCTTTCAGAGAGATTTTATAGA 1558  
Qy 1441 GAGATTTCTATTAATTTTGTATACAGTGTGATGCTGAACTCTTTCAGAGAGATTTTATAGA 1500  
Db 1559 GAGATTTCTATTAATTTTGTATACAGTGTGATGCTGAACTCTTTCAGAGATTTTATAGA 1618  
Qy 1501 AAATCATTTGAATTC 1514  
Db 1619 AAATCATTTGAATTC 1632

RESULT 4  
BD103190  
LOCUS BD103190 11284 bp DNA linear PAT 27-AUG-2002  
DEFINITION O-fucosyltransferase.  
ACCESSION BD103190  
VERSION BD103190.1 GI:22648764  
KEYWORDS JP 2001527389-A/3.  
SOURCE unidentified  
ORGANISM unclassified  
REFERENCE 1 (bases 1 to 11284)  
AUTHORS Wang, Y. and Spellman, M. W.  
TITLE O-fucosyltransferase.  
JOURNAL Patent JP 2001527389-A 3 25-DEC-2001;  
COMMENT GENENTECH INC  
OS Unidentified  
PN JP 2001527389-A/3  
PD 25-DEC-2001  
PF 17-DEC-1997 JP 1998532877 08/978741 PI  
PR 31-JAN-1997 US 08/752498, 26-NOV-1997 US  
PC C12N15/54, C12N9/10, C07K16/40  
CC Strandedness: Single;  
CC Topology: Linear;  
CC O-fucosyltransferase  
FT Key Location/Qualifiers  
FT source 1..11284 /organism='Unidentified'.  
FEATURES  
source Location/Qualifiers  
1..11284 /organism='unidentified'  
/mol\_type='genomic DNA'  
/db\_xref='taxon:32644'

ORIGIN  
Query Match 99.9%; Score 1512.4; DB 6; Length 11284;  
Best Local Similarity 99.9%; Pred No. 0;  
Matches 1513; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 ATGCCCGGGCTCTCTGGGACCGGCGGGTTACTGTCTACTGCGCCCTGATCGGGCGC 60  
Db 4236 ATGCCCGGGCTCTCTGGGACCGGCGGGTTACTGTCTACTGCGCCCTGATCGGGCGC 4295

```
Qy 61 TTTGGGAACCAAGCGATCACTTTCTGGGCTCTCTGGCATTTGCAAGCTGCTAAACCGT 120
Db |||||
4296 TTTGGGAACCAAGCGCGATCACTTTCTGGCTCTCTGGCATTTGCAAGCTGCTAAACCGT 4355
Qy 121 ACCTTGGCTGTCCTCTCTTGGATTTAGTACAGCATCAAGAGCTCTCTTTCACCAACCTC 180
Db |||||
4356 ACCTTGGCTGTCCTCTCTTGGATTTAGTACAGCATCAAGAGCTCTCTTTCACCAACCTC 4415
Qy 181 CATGTGCTCTACCAAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGGTCATC 240
Db |||||
4416 CATGTGCTCTACCAAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGGTCATC 4475
Qy 241 AGCTTGGAGGATTTTCATGGAGAAGCTGGCACCCACCTGCGCCCTTGAGAAGCGGGTG 300
Db |||||
4476 AGCTTGGAGGATTTTCATGGAGAAGCTGGCACCCACCTGCGCCCTTGAGAAGCGGGTG 4535
Qy 301 GCATACCTCTTGGAGTGGACCCAGCGAAGCCAGATAGAGACGTGCCCATGAAG 360
Db |||||
4536 GCATACCTCTTGGAGTGGACCCAGCGAAGCCAGATAGAGACGTGCCCATGAAG 4595
Qy 361 GAAGAAACCCCTTTGGGCCATTTCTGGGATCAGTTTTCATGTGAGTTTCAACAAGTCGGAG 420
Db |||||
4596 GAAGAAACCCCTTTGGGCCATTTCTGGGATCAGTTTTCATGTGAGTTTCAACAAGTCGGAG 4655
Qy 421 CTTTTCACAGGATTTCTCTCAGTGCTTCTACAGAGAACAATGGAGCCAGAGATTTTCT 480
Db |||||
4656 CTTTTCACAGGATTTCTCTCAGTGCTTCTACAGAGAACAATGGAGCCAGAGATTTTCT 4715
Qy 481 CCAAGGAACATCGGCTGCTGCTGCGAGAGCCCGAGCCAGCTTCCCGTCTAGAA 540
Db |||||
4716 CCAAGGAACATCGGCTGCTGCTGCGAGAGCCCGAGCCAGCTTCCCGTCTAGAG 4775
Qy 541 GAAACACGCGCATACAGAAGTACATGTATGCTCAGACGAATGCTGAACACGGGAGAG 600
Db |||||
4776 GAAACACGCGCATACAGAAGTACATGTATGCTCAGACGAATGCTGAACACGGGAGAG 4835
Qy 601 GCCCAGATTCATGCCCATCTTGTGCGGCCCTATGTGGGCATTCATCTGCGCATTTGGCTCT 660
Db |||||
4836 GCCCAGATTCATGCCCATCTTGTGCGGCCCTATGTGGGCATTCATCTGCGCATTTGGCTCT 4895
Qy 661 GACTGGAAGAACGCTGTGCGCATGTGAAGACGGGATGCGAGGCTGCGCATTTGATGCC 720
Db |||||
4896 GACTGGAAGAACGCTGTGCGCATGTGAAGACGGGATGCGAGGCTGCGCATTTGATGCC 4955
Qy 721 TCTCGCAGTGTGGGCTACAGCGCAGCAGCGGCCCTCAGCATCACTATGTC 780
Db |||||
4956 TCTCGCAGTGTGGGCTACAGCGCAGCAGCGGCCCTCAGCATCACTATGTC 5015
Qy 781 CTGCTGACCTGAAGGATCCAGAGGCTGTGAAGCTCTGGGTGAGTCTGCTGGATGCC 840
Db |||||
5016 CTGCTGACCTGAAGGATCCAGAGGCTGTGAAGCTCTGGGTGAGTCTGCTGGATGCC 5075
Qy 841 CAGTCGCTGCTAGTGTCTGATTTCCAGAGTTATGTGCTCTGAGTCTCAACAGCTCTTC 900
Db |||||
5076 CAGTCGCTGCTAGTGTCTGATTTCCAGAGTTATGTGCTCTGAGTCTCAACAGCTCTTC 5135
Qy 901 AAAGGAAGTGAAGTGTGAGCTGAAGCTGAGGTGCGCCAGGTCGACCTGTACATC 960
Db |||||
5136 AAAGGAAGTGAAGTGTGAGCTGAAGCTGAGGTGCGCCAGGTCGACCTGTACATC 5195
Qy 961 CTCGGCCAAAGCCGACCTTTATTGGCAACTGTGTCTCTCTTCACTGCTCTTGTGAAG 1020
Db |||||
5196 CTCGGCCAAAGCCGACCTTTATTGGCAACTGTGTCTCTCTTCACTGCTCTTGTGAAG 5255
Qy 1021 CGGAGCGGACCTTCAGGGAGGCGGCTCTTTCTTTCGGGATGGACAGGCCCCCTTAAG 1080
Db |||||
5256 CGGAGCGGACCTTCAGGGAGGCGGCTCTTTCTTTCGGGATGGACAGGCCCCCTTAAG 5315
Qy 1081 CTGGGAGCAGTGTCTGATTTCTGGCCGAGCAGACCTCTGATCTGAGGGGACGAG 1140
Db |||||
5316 CTGGGAGCAGTGTCTGATTTCTGGCCGAGCAGACCTCTGATCTGAGGGGACGAG 5375
```

```
Qy 1141 AGTCTGAGCTGCTCTTTCAGCCAGGCGCTGCGACGAGGAGTGTCTCCGGGATTTGCAAACT 1200
Db |||||
5376 AGTCTGAGCTGCTCTTTCAGCCAGGCGCTGCGACGAGGAGTGTCTCCGGGATTTGCAAACT 5435
Qy 1201 CCTCTTCTCACTGCCCAAGATGGAGAAGTGGAGGACCCCTCAAGGAGGGAGGAGCGC 1260
Db |||||
5436 CCTCTTCTCACTGCCCAAGATGGAGAAGTGGAGGACCCCTCAAGGAGGGAGGAGCGC 5495
Qy 1261 TCCATATCCCAAGGCGATAGGACTTTCAGGTTCTTAGGAGCAGGAGCATCTCCCATCGCAC 1320
Db |||||
5496 TCCATATCCCAAGGCGATAGGACTTTCAGGTTCTTAGGAGCAGGAGCATCTCCCATCGCAC 5555
Qy 1321 GTGCTTTCTGCTCTTCTGGGAAATTTCTCACACTGGCAAGCAGTCCAGCTCCGCTTCT 1380
Db |||||
5556 GTGCTTTCTGCTCTTCTGGGAAATTTCTCACACTGGCAAGCAGTCCAGCTCCGCTTCT 5615
Qy 1381 GGTCACCTCTGCTGAGCAGCCTGGGATGCTGAATCTTTCAGAGAGATTTTATATAGA 1440
Db |||||
5616 GGTCACCTCTGCTGAGCAGCCTGGGATGCTGAATCTTTCAGAGAGATTTTATATAGA 5675
Qy 1441 GAGATTTCTATAATTTTGATACAAAGTTCATGACTATCTAGAACTCTCTGTGGTTTTTGA 1500
Db |||||
5676 GAGATTTCTATAATTTTGATACAAAGTTCATGACTATCTAGAACTCTCTGTGGTTTTTGA 5735
Qy 1501 AAATCATTTGAATTC 1514
Db |||||
5736 AAATCATTTGAATTC 5749

RESULT 5
LOCUS D80002 5189 bp mRNA linear PRI 10-JAN-2003
DEFINITION Homo sapiens KIAA0180 mRNA, complete cds.
ACCESSION D80002
VERSION D80002.2 GI:20521837
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 Nagase,T., Seki,N., Iehikawa,K., Tanaka,A. and Nomura,N.
Prediction of the coding sequences of unidentified human genes. V.
The coding sequences of 40 new genes (KIAA0161-KIAA0200) deduced by
analysis of cDNA clones from human cell line KG-1
DNA Res. 3 (1), 17-24 (1996)
JOURNAL 96281124
MEDLINE 8724849
REFERENCE
2 Chiang,P.W., Wang,S., Smithivas,P., Song,W.J., Ramamoorthy,S.,
Hillman,J., Puett,S., Van Keuren,M.L., Crombez,E., Kumar,A.,
Glover,T.W., Miller,D.E., Tsai,C.H., Blackburn,C.C., Chen,X.N.,
Sun,Z., Cheng,J.F., Korenberg,J.R. and Kurnit,D.M.
Identification and analysis of the human and murine putative
chromatin structure regulator SUPT6H and Supt6h
Genomics 34 (3), 328-333 (1996)
JOURNAL 96374824
MEDLINE 8786132
REFERENCE
3 (bases 1 to 5189)
Ohara,O., Nagase,T., Kikuno,R. and Nomura,N.
Direct Submission
Submitted (12-DEC-1995) Osamu Ohara, Kazusa DNA Research Institute;
1532-3, Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail:cdnainfo@kazusa.or.jp, Tel:+81-438-52-3913)
On May 9, 2002 this sequence version replaced gi:1136419.
COMMENT
FEATURES
Location/Qualifiers
1..5189
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="ha0256781"
/sex="male"
/cell_line="KG-1"
```



AUTHORS Haferlach, T., Schoch, C., Kern, W., Kohlmann, A., Schnittger, S., Dugas, M., Eils, R., Brose, B. and Mergenthaler, S.

TITLE Novel genetic markers for leukemias

JOURNAL Patent: WO 03039443-A 2310 15-MAY-2003; Deutsches Krebsforschungszentrum (DE) ; Ludwig-Maximilian-Universitaet Muenchen (DE) ; PD Dr. Dr. (DE) ; Schoch, Claudia (DE) ; Kern, Wolfgang (DE)

FEATURES

source

1..5266

Location/Qualifiers

/organism="Homo sapiens"

/mol\_type="unassigned DNA"

/db\_xref="taxon:9606"

Query Match	99.7%	Score 1509.8	DB 6	Length 5266
Best Local Similarity	99.8%	Pred. No. 0		
Matches 1511	Conservative 0	Mismatches 3	Indels 0	Gaps 0
QY	1	ATGCCGCGGGCTCTCGGACCCGCGGTTACCTGCTCTACTGCCCTCATGGGCGC	60	
Db	132	ATGCCCTGGGGTCTCTGGGACCCGCGCGGTTACCTGCTCTACTGCCCTCATGGGCGC	191	
QY	61	TTTGGGAACACGCGCGATCACTTCTTGGGCTCTCTGGCATTTGCAAGCTGCTAAACCGT	120	
Db	192	TTTGGGAACACGCGCGATCACTTCTTGGGCTCTCTGGCATTTGCAAGCTGCTAAACCGT	251	
QY	121	ACCTTGGGTGTCCTCTTGGATTGAGTACACAGCATCACAAAGCTCTCTTTACCAACCTTC	180	
Db	252	ACCTTGGGTGTCCTCTTGGATTGAGTACACAGCATCACAAAGCTCTCTTTACCAACCTTC	311	
QY	181	CATGTGTCTTACCAAGTACTTCAAGCTGGAGCCCCTCAGGCTTTACCATCGGGTCATC	240	
Db	312	CATGTGTCTTACCAAGTACTTCAAGCTGGAGCCCCTCAGGCTTTACCATCGGGTCATC	371	
QY	241	AGCTTGGAGGATTTCAATCGGAAGCTGSCACCCACCTGCGCCCTCGAAGACGGGTG	300	
Db	372	AGCTTGGAGGATTTCAATCGGAAGCTGSCACCCACCTGCGCCCTCGAAGACGGGTG	431	
QY	301	GCATACTGCTTTGAGTGGCAGCCAGCGAAGCCAGATAAGAAAGAGCTGCCCCCATGAAG	360	
Db	432	GCATACTGCTTTGAGTGGCAGCCAGCGAAGCCAGATAAGAAAGAGCTGCCCCCATGAAG	491	
QY	361	GAAGGAAACCCCTTTGGGCCATTCCTGGGATCAGTTTTCATGTGAGTTTCAACAAGTCGAG	420	
Db	492	GAAGGAAACCCCTTTGGGCCATTCCTGGGATCAGTTTTCATGTGAGTTTCAACAAGTCGAG	551	
QY	421	CTTTTTCACAGGATTTTCCTTCAGTGCTCTCTACAGAAACAATGAGGACGAGATTTTCT	480	
Db	552	CTTTTTCACAGGATTTTCCTTCAGTGCTCTCTACAGAAACAATGAGGACGAGATTTTCT	611	
QY	481	CCAAAGGAACATCCGGTCTTGCCTCGCAGGAGCCCGCCAGTTTCCCGTCTTAGAA	540	
Db	612	CCAAAGGAACATCCGGTCTTGCCTCGCAGGAGCCCGCCAGTTTCCCGTCTTAGAA	671	
QY	541	GAACACAGGCCACTACAGAAGTACATGGTATGGTTCAGACGAAATGGTGAAGACGGGAG	600	
Db	672	GAACACAGGCCACTACAGAAGTACATGGTATGGTTCAGACGAAATGGTGAAGACGGGAG	731	
QY	601	GCCAGATTATGCCCACTTGTGCGGCCCTATGTGGGCATTCATCTGGCATTTGCTCT	660	
Db	732	GCCAGATTATGCCCACTTGTGCGGCCCTATGTGGGCATTCATCTGGCATTTGCTCT	791	
QY	661	GACTCGAAGACGCTGTGCCATGCTGAAGACGGGACTCGAGGCTCGCATTTATGGCC	720	
Db	792	GACTCGAAGACGCTGTGCCATGCTGAAGACGGGACTCGAGGCTCGCATTTATGGCC	851	
QY	721	TCTCCGCAAGTGTGGGCTACAGCGGACAGACAGCGGCCCTCTCACATGACTATGTGC	780	
Db	852	TCTCCGCAAGTGTGGGCTACAGCGGACAGACAGCGGCCCTCTCACATGACTATGTGC	911	
QY	781	CTGCTGACCTGAAGGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGGTCGTGGATGCC	840	
Db	912	CTGCTGACCTGAAGGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGGTCGTGGATGCC	971	





Query Match		95.6%;	Score 1447.4;	DB 6;	Length 5009;	
Best Local Similarity		99.9%;	Pred. No. 0;			
Matches 1448;		Conservative	0;	Mismatches	1;	Indels 0; Gaps 0;
QY	66	GAACAGGCGGATCACTTCTTGGGCTCTCTGGCATTTGGCAAGTGTGCAAAAGCTGTAAACCGTACCTTT	125			
DB	1	GAACAGGCGGATCACTTCTTGGGCTCTCTGGCATTTGGCAAGTGTGCAAAAGCTGTAAACCGTACCTTT	60			
QY	126	GGCTGTCCCTCTTGGATTGAGTACAGCATCACAGGCTCTCTTTCACCAACCTCCATGT	185			
DB	61	GGCTGTCCCTCTTGGATTGAGTACAGCATCACAGGCTCTCTTTCACCAACCTCCATGT	120			
QY	186	GTCTACCAAGAGTACTTCAAGCTGGAGCCCTTCCAGGCTTACCATCGGGTCAATCAGCTT	245			
DB	121	GTCTACCAAGAGTACTTCAAGCTGGAGCCCTTCCAGGCTTACCATCGGGTCAATCAGCTT	180			
QY	246	GGAGGATTTTCAGGAGAGCTGGCACCACCCACTGGCCCCCTTGAGAGCGGGTGGCATA	305			
DB	181	GGAGGATTTTCAGGAGAGCTGGCACCACCCACTGGCCCCCTTGAGAGCGGGTGGCATA	240			
QY	306	CTGCTTTGAGGTGGCAGCCAGCGAGCCAGATGAAGAGCGTGCCTCATGAAGGAGG	365			
DB	241	CTGCTTTGAGGTGGCAGCCAGCGAGCCAGATGAAGAGCGTGCCTCATGAAGGAGG	300			
QY	366	AAACCCCTTTGGCCCATCTGGGATCAGTTTTCATGTGAGTTTCAACAAGTCGGAGCTTTT	425			
DB	301	AAACCCCTTTGGCCCATCTGGGATCAGTTTTCATGTGAGTTTCAACAAGTCGGAGCTTTT	360			
QY	426	TACAGGCATTTCTTCAGTGCTTCTACAGAGAACATGGAGCCAGAGATTTTCTCCAAA	485			
DB	361	TACAGGCATTTCTTCAGTGCTTCTACAGAGAACATGGAGCCAGAGATTTTCTCCAAA	420			
QY	486	GAACATCCGCTGTTCCTGCCAGGAGCCAGCCAGTTCCTCCGCTCCTAGAGAGACA	545			
DB	421	GAACATCCGCTGTTCCTGCCAGGAGCCAGCCAGTTCCTCCGCTCCTAGAGAGACA	480			
QY	546	CAGGCCACTACAGAGTACATGATGTGTGTCAGACGAAATGTGAAGACGGAGAGGCCCA	605			
DB	481	CAGGCCACTACAGAGTACATGATGTGTGTCAGACGAAATGTGAAGACGGAGAGGCCCA	540			
QY	606	GATTCATGCCACCTTGTCCGCCCTATGTGGGCATTCATCTGGCGATTTGGCTGTGACTG	665			
DB	541	GATTCATGCCACCTTGTCCGCCCTATGTGGGCATTCATCTGGCGATTTGGCTGTGACTG	600			
QY	666	GAAGAACGCTGTGCCATGCTCAAGACGGGACTCGAGCTCGCACTTCATGCGCTCTCC	725			
DB	601	GAAGAACGCTGTGCCATGCTCAAGACGGGACTCGAGCTCGCACTTCATGCGCTCTCC	660			
QY	726	GCAGTGTGTGGCTACAGCCGACAGCAGCGGCCCTCCCTCAGATGACTATGTGCTGCC	785			
DB	661	GCAGTGTGTGGCTACAGCCGACAGCAGCGGCCCTCCCTCAGATGACTATGTGCTGCC	720			
QY	786	TGACCTGAAGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGGTGCTGGATGCCAGTTC	845			
DB	721	TGACCTGAAGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGGTGCTGGATGCCAGTTC	780			
QY	846	GGTCTACGTGTCTACTGATTCGAGAGTTATGTGCTGAGCTCCACAGCTCTTCAAAGG	905			
DB	781	GGTCTACGTGTCTACTGATTCGAGAGTTATGTGCTGAGCTCCACAGCTCTTCAAAGG	840			
QY	906	GAAGGTGAAGTGTGAGCTGAAGCTGAGGTGGCCAGGTCGACCTGTATCATCTCTGG	965			
DB	841	GAAGGTGAAGTGTGAGCTGAAGCTGAGGTGGCCAGGTCGACCTGTATCATCTCTGG	900			
QY	966	CCAAGCCGACCACTTTATGGCAACTGTGTCTCTCTTCACTGCTTTGTGAAGCGGA	1025			
DB	901	CCAAGCCGACCACTTTATGGCAACTGTGTCTCTCTTCACTGCTTTGTGAAGCGGA	960			
QY	1026	CGGGACCTCCAGGGGAGCGCTCTCTTCTTGGCATGGACAGGCCCTTAAGCTGG	1085			
DB	961	CGGGACCTCCAGGGGAGCGCGTCTCTTCTTTCGGCATGGACAGGCCCTTAAGCTGG	1020			
QY	1086	GGACGAGTTCTGATTTCTGGCCGGAGCACAGACCCCTCTGATCTCTGGAGGACGAGTCT	1145			

DB	1021	GGACGAGTTCTGATTTCTGGCCGAGCACCAGACCCCTCTGATCTCTGGAGGACGAGTCT	1080			
QY	1146	GAGCTGGTCTCTTCCAGCCAGGCTGGCAGCCAGAGTGTCTCGGATTTGCAAACTCTCT	1205			
DB	1081	GAGCTGGTCTCTTCCAGCCAGGCTGGCAGCCAGAGTGTCTCGGATTTGCAAACTCTCT	1140			
QY	1206	TCTCACCTGGCCAAAGATGGAGAGAGTGCAGGGACCCCTCAAGGAGGAGAGCGTCCAT	1265			
DB	1141	TCTCACCTGGCCAAAGATGGAGAGAGTGCAGGGACCCCTCAAGGAGGAGAGCGTCCAT	1200			
QY	1266	ATCCAGGCGCATAGGACTTTGCAGTTCCTTAGGAGCAGGAGCATCTCCCATCGAGTGT	1325			
DB	1201	ATCCAGGCGCATAGGACTTTGCAGTTCCTTAGGAGCAGGAGCATCTCCCATCGAGTGT	1260			
QY	1326	TTCCTCTCTTCTGGGAATTTCTCACCTGGCCAAAGCAGTCCAGCTCTCGTCTTCTGGTCC	1385			
DB	1261	TTCCTCTCTTCTGGGAATTTCTCACCTGGCCAAAGCAGTCCAGCTCTCGTCTTCTGGTCC	1320			
QY	1386	ACTCTGCTCTCAGCAGCCTGGGATGCTGAACCTCTTTCAGAGAGATTTTATAGAGAT	1445			
DB	1321	ACTCTGCTCTCAGCAGCCTGGGATGCTGAACCTCTTTCAGAGAGATTTTATAGAGAT	1380			
QY	1446	TCTATAATTTTGATACAAAGTTCATGACTATCTCTAGAACTCTCTCTGTTTGAATAATC	1505			
DB	1381	TCTATAATTTTGATACAAAGTTCATGACTATCTCTAGAACTCTCTCTGTTTGAATAATC	1440			
QY	1506	ATTGAATTC 1514				
DB	1441	ATTGAATTC 1449				
RESULT 9						
LOCUS	BD103191	O-fucosyltransferase.	5009 bp	DNA	linear	PAT 27-AUG-2002
DEFINITION	BD103191	O-fucosyltransferase.				
ACCESSION	BD103191					
VERSION	BD103191.1	GI:22648765				
KEYWORDS	JP 2001527389-A/4.					
SOURCE	unidentified					
ORGANISM	unclassified					
REFERENCE	1 (bases 1 to 5009)					
AUTHORS	Wang, Y. and Spellman, M.W.					
TITLE	O-fucosyltransferase					
JOURNAL	Patent: JP 2001527389-A 4 25-DEC-2001;					
COMMENT	GENENTECH INC					
	OS Unidentified					
	PN JP 2001527389-A/4					
	PD 25-DEC-2001					
	PF 17-DEC-1997 JP 1998532877					
	PR 31-JAN-1997 US 08/792498, 26-NOV-1997 US 08/978741 PI					
	YANG WANG, MICHAEL W SPELLMAN					
	PC C12N15/54, C12N9/10, C07K16/40					
	CC Strandedness: Single;					
	CC Topology: Linear;					
	CC O-fucosyltransferase					
	FT Key	Location/Qualifiers				
	FT source	1..5009				
	FT	/organism='Unidentified'.				
FEATURES						
	source	1..5009				
		/organism="unidentified"				
		/mol_type="genomic DNA"				
		/db_xref="taxon:32644"				
ORIGIN						
Query Match						
Best Local Similarity						
Matches 1448; Conservative						
0; Mismatches 0; Indels 0; Gaps 0;						
QY	66	GAACAGGCGGATCACTTCTTGGGCTCTCTGGCATTTGGCAAGTGTGCAAAAGCTGTAAACCGTACCTTT	125			

Db 1 GAACAGCCGAGTACATCTTTGGGCTCTCTGGCATTTGCAAGCTGCTAAACCGTACCTT 60  
Qy 126 GGCTGTCCCTCTTTGGATTGAGTACAGCATCAAAAGCTCTCTTTTCAACAACTCCATGT 185  
Db 61 GGCTGTCCCTCTTTGGATTGAGTACAGCATCAAAAGCTCTCTTTTCAACAACTCCATGT 120  
Qy 186 GTCTTACAGAACTACTTCAAGCTGAGCCCTCCAGGCTTACATCGGGTCACTAGCTT 245  
Db 121 GTCTTACAGAACTACTTCAAGCTGAGCCCTCCAGGCTTACATCGGGTCACTAGCTT 180  
Qy 246 GGAGGATTTATGAGAACTGGCAACCACTGAGCTGAGCCCTGAGAGCGGGTGGATA 305  
Db 181 GGAGGATTTATGAGAACTGGCAACCACTGAGCTGAGCCCTGAGAGCGGGTGGATA 240  
Qy 306 CTGCTTTGAGTGGCAGCCAGCAAGCCGAGATAGAGACGTGCCCCATGAGGAAGG 365  
Db 241 CTGCTTTGAGTGGCAGCCAGCAAGCCGAGATAGAGACGTGCCCCATGAGGAAGG 300  
Qy 366 AAACCCCTTTGGCCATTCFGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAGCTTTT 425  
Db 301 AAACCCCTTTGGCCATTCFGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAGCTTTT 360  
Qy 426 TACAGGCATTTCTTCAAGTCTTCTTACAGAGAAACAATGAGGCACAGATTTTCTCCAAA 485  
Db 361 TACAGGCATTTCTTCAAGTCTTCTTACAGAGAAACAATGAGGCACAGATTTTCTCCAAA 420  
Qy 486 GGAAATCCCGTGTCTGCTGTCAGAGAGCCCGAGCCAGTTCCTCCGCTCTTAGAAGAA 545  
Db 421 GGAAATCCCGTGTCTGCTGTCAGAGAGCCCGAGCCAGTTCCTCCGCTCTTAGAAGAA 480  
Qy 546 CAGGCCACTACAGAACTACATGTTATGTCAGAGAAATGTCAGAGCGGAGAGGCCA 605  
Db 481 CAGGCCACTACAGAACTACATGTTATGTCAGAGAAATGTCAGAGCGGAGAGGCCA 540  
Qy 606 GATTCATGCCCCCTGTTGGCCCTGATGTCGGAATTCATCTCGCATGCGTCTGACTG 665  
Db 541 GATTCATGCCCCCTGTTGGCCCTGATGTCGGAATTCATCTCGCATGCGTCTGACTG 600  
Qy 666 GAAGAACCTCTGCCCCTGTCAGAGAGCGGACTGCGAGCTTCAATGAGCTCTTCC 725  
Db 601 GAAGAACCTCTGCCCCTGTCAGAGAGCGGACTGCGAGCTTCAATGAGCTCTTCC 660  
Qy 726 GCAGTGTGGGCTACAGCCGAGAGACAGCGGCCCTTCAAGATGATGTCCTGCC 785  
Db 661 GCAGTGTGGGCTACAGCCGAGAGACAGCGGCCCTTCAAGATGATGTCCTGCC 720  
Qy 786 TGACCTGAAGGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGGTGCGTGGATGCCAGTC 845  
Db 721 TGACCTGAAGGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGGTGCGTGGATGCCAGTC 780  
Qy 846 GGTCTACGTTGCTACTGATTCGAGAGTTATGTGCTGAGCTCAACAGCTCTTCAAAGG 905  
Db 781 GGTCTACGTTGCTACTGATTCGAGAGTTATGTGCTGAGCTCAACAGCTCTTCAAAGG 840  
Qy 906 GAAGTGAAGGTGTGAGCTGAAGCTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAG 965  
Db 841 GAAGTGAAGGTGTGAGCTGAAGCTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAG 900  
Qy 966 CCAAGCGAGACCTTTATTTGGGCAATGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1025  
Db 901 CCAAGCGAGACCTTTATTTGGGCAATGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 960  
Qy 1026 GCGGAGCTTCAAGGGAGGCGCTCTTCTTCTTGGGATGGAAGAGCCCTTAAAGTGGG 1085  
Db 961 GCGGAGCTTCAAGGGAGGCGCTCTTCTTCTTGGGATGGAAGAGCCCTTAAAGTGGG 1020  
Qy 1086 GGAAGAGTTCTGATTTGGCCGAGAGACAGAGCCCTCTGATCTTGGAGGACAGAGTCT 1145  
Db 1021 GGAAGAGTTCTGATTTGGCCGAGAGACAGAGCCCTCTGATCTTGGAGGACAGAGTCT 1080  
Qy 1146 GAGCTGTGCTTCCAGCCAGGCTTGGCAGCCAGAGGCTCTCCGGGATTCGAACTCTCTCT 1205  
Db 1081 GAGCTGTGCTTCCAGCCAGGCTTGGCAGCCAGAGGCTCTCCGGGATTCGAACTCTCTCT 1140

Qy 1206 TCTCACCTGCCAAAGATGGAGAGAGTGCAGGAGACCCCTCAAGGAGGAGAGCGCTCCAT 1265  
Db 1141 TCTCACCTGCCAAAGATGGAGAGAGTGCAGGAGACCCCTCAAGGAGGAGAGCGCTCCAT 1200  
Qy 1266 ATCCAGAGGATAGGACTTGCAGGTTCTTAGGAGCAGAGCATCTCCCATCGACGTGCT 1325  
Db 1201 ATCCAGAGGATAGGACTTGCAGGTTCTTAGGAGCAGAGCATCTCCCATCGACGTGCT 1260  
Qy 1326 TTCTGCTCTTCTGGGAATTTCTCACTGGCAAAAGCAGTCCAGCCTCGCTCTTCTGCTCC 1385  
Db 1261 TTCTGCTCTTCTGGGAATTTCTCACTGGCAAAAGCAGTCCAGCCTCGCTCTTCTGCTCC 1320  
Qy 1386 ACTCTGCTCTGAGAGCGCTGGGATGCTGAACTCTTTCAGAGAGATTTTTTTATAGAGAT 1445  
Db 1321 ACTCTGCTCTGAGAGCGCTGGGATGCTGAACTCTTTCAGAGAGATTTTTTTATAGAGAT 1380  
Qy 1446 TTCTATAATTTTGATACAAAGTTCATGACTATCTCTAGAACTCTCTGTGGTTTTTGAATC 1505  
Db 1381 TTCTATAATTTTGATACAAAGTTCATGACTATCTCTAGAACTCTCTGTGGTTTTTGAATC 1440  
Qy 1506 ATTGAATTC 1514  
Db 1441 ATTGAATTC 1449

RESULT 10  
AR163452  
LOCUS AR163452 1300 bp DNA linear PAT 17-OCT-2001  
DEFINITION Sequence 6 from patent US 6270987.  
ACCESSION AR163452  
VERSION AR163452.1 GI:16234057  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE Unclassified.  
1 (bases 1 to 1300)  
AUTHORS Wang, Y. and Spellman, M. W.  
TITLE O-fucosyltransferase  
JOURNAL Patent: US 6270987-A 6 07-AUG-2001;  
FEATURES Location/Qualifiers  
source 1..1300  
/organism="unknown"  
/mol\_type="unassigned DNA"

ORIGIN

Query Match 76.8%; Score 1163.4; DB 6; Length 1300;  
Best Local Similarity 99.9%; Pred. No. 2.7e-284;  
Matches 1164; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATGCCCGCGGGCTCTTGGGACCCGCGGTTACCTGCTTACTGCCCCCTGCATGGGCGC 60  
Db 136 ATGCCCGCGGGCTCTTGGGACCCGCGGTTACCTGCTTACTGCCCCCTGCATGGGCGC 195  
Qy 61 TTTGGGAACCAAGCGCGATCACTTTCTTGGGCTCTCTGCGATTTGCAAGCTGCTAAACCGT 120  
Db 196 TTTGGGAACCAAGCGCGATCACTTTCTTGGGCTCTCTGCGATTTGCAAGCTGCTAAACCGT 255  
Qy 121 ACCTTGGCTGTCCCTCTGGATGAGTACAGGATCAGAGCTGCTTTCACCAAGCTC 180  
Db 256 ACCTTGGCTGTCCCTCTGGATGAGTACAGGATCAGAGCTGCTTTCACCAAGCTC 315  
Qy 181 CATGTGCTCTTACCAAGATCACTTCAAGCTGAGGCCCTCCAGGCTTACCATGGGTCAATC 240  
Db 316 CATGTGCTCTTACCAAGATCACTTCAAGCTGAGGCCCTCCAGGCTTACCATGGGTCAATC 375  
Qy 241 AGCTTGGAGGATTTTCAAGAGAGTGGCAACCACTGGCCCCCTGAGAGCGGGTG 300  
Db 376 AGCTTGGAGGATTTTCAAGAGAGTGGCAACCACTGGCCCCCTGAGAGCGGGTG 435  
Qy 301 GCATACCTCTTTGAGGTTGGCAGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 360  
Db 436 GCATACCTCTTTGAGGTTGGCAGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 495



```
QY 361 GAAGGAAACCCCTTTGGGCCAATTCGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
Db 496 GAAGGAAACCCCTTTGGGCCAATTCGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 555
QY 421 CTTTATTACAGGATTTCTTTCATGCTTCTTCAAGAGAAATGAGGACGAGATTTTCT 480
Db 556 CTTTATTACAGGATTTCTTTCATGCTTCTTCAAGAGAAATGAGGACGAGATTTTCT 615
QY 481 CCAAGGAAACATCCGGTCTTCCCTGCCAGAGCCCGAGCCAGTTCCTCCGTCCTAGNA 540
Db 616 CCAAGGAAACATCCGGTCTTCCCTGCCAGAGCCCGAGCCAGTTCCTCCGTCCTAGAG 675
QY 541 GAACACAGGCCCACTACAGAAGTACATGTTGCTCAGACGAATGTTGAAGACGGGAGAG 600
Db 676 GAACACAGGCCCACTACAGAAGTACATGTTGCTCAGACGAATGTTGAAGACGGGAGAG 735
QY 601 GCCAGATTTCATGCCACCTTGTCCGGCCCTATGTGGGCATTTCATCTGCGCATTCGGCTCT 660
Db 736 GCCAGATTTCATGCCACCTTGTCCGGCCCTATGTGGGCATTTCATCTGCGCATTCGGCTCT 795
QY 661 GACTCGAAGACGCTGTGCCATGCTGAAGGACGGGACTGACAGGCTCGCACTTCATGGCC 720
Db 796 GACTCGAAGACGCTGTGCCATGCTGAAGGACGGGACTGACAGGCTCGCACTTCATGGCC 855
QY 721 TCTCCGCACTGTGTGGGCTACAGCCGACAGCAGCGCCGCCCTCACGATGACTATGTGC 780
Db 856 TCTCCGCACTGTGTGGGCTACAGCCGACAGCAGCGCCGCCCTCACGATGACTATGTGC 915
QY 781 CTGCTGACCTGAAGGAGATCAGAGGGCTGTGAAGCTCTGGGTGAGGTTCGCTGGATGCC 840
Db 916 CTGCTGACCTGAAGGAGATCAGAGGGCTGTGAAGCTCTGGGTGAGGTTCGCTGGATGCC 975
QY 841 CAGTCGGTCTAGTTGCTACTGATTCGAGAGTTATGTGCTGAGCTCCAGCTCCACAGCTCTTC 900
Db 976 CAGTCGGTCTAGTTGCTACTGATTCGAGAGTTATGTGCTGAGCTCCACAGCTCTTC 1035
QY 901 AAAGGGAAGGTGAAGGTGTGAGCTGAAGCTGAGGTGGCCGAGCTCGACTGTACATC 960
Db 1036 AAAGGGAAGGTGAAGGTGTGAGCTGAAGCTGAGGTGGCCGAGCTCGACTGTACATC 1095
QY 961 CTCGCCCAAGCCGACACTTTATTTGGCAACTGTGTCTCTCTTCACTGCTTTGTGAAG 1020
Db 1096 CTCGCCCAAGCCGACACTTTATTTGGCAACTGTGTCTCTTCACTGCTTTGTGAAG 1155
QY 1021 CGGAGCGGACCTCCAGGGAGGCGCTCTTCTTCTCGGATGGAAGCCGCTTAAG 1080
Db 1156 CGGAGCGGACCTCCAGGGAGGCGCTCTTCTTCTCGGATGGAAGCCGCTTAAG 1215
QY 1081 CTGCGGACGAGTTCTGATTTCTGGCGGAGCACCAGACCTCTGATCTGGAGGACCCAG 1140
Db 1216 CTGCGGACGAGTTCTGATTTCTGGCGGAGCACCAGACCTCTGATCTGGAGGACCCAG 1275
QY 1141 AGTCTGAGCTGCTCTTCAGCCAG 1165
Db 1276 AGTCTGAGCTGCTCTTCAGCCAG 1300
```

```
RESULT 11
BD103189
LOCUS BD103189 1300 bp DNA linear PAT 27-AUG-2002
DEFINITION O-fucosyltransferase.
ACCESSION BD103189
VERSION BD103189.1 GI:22648763
KEYWORDS JP 2001527389-A/2.
SOURCE unclassified
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 1300)
AUTHORS Wang, Y. and Spellman, M. W.
TITLE O-fucosyltransferase
JOURNAL Patent: JP 2001527389-A 2 25-DEC-2001;
GENENTECH INC
```

```
COMMENT OS Unidentified
PN JP 2001527389-A/2
PD 25-DEC-2001
PF 17-DEC-1997 JP 1998532877
PR 31-JAN-1997 US 08/792498, 26-NOV-1997 US 08/978741 PI
YANG WANG, MICHAEL W SPELLMAN
C12N15/54, C12N9/10, C07K16/40
PC Strandedness: Single;
CC Topology: linear;
CC O-fucosyltransferase
FH Key Location/Qualifiers
FT source 1..1300
FT /organism='Unidentified'.
FEATURES
source
1..1300
/organism='unidentified'
/mol_type='genomic DNA'
/db_xref='taxon:32644'
ORIGIN
Query Match 76.8%; Score 1163.4; DB 6; Length 1300;
Best Local Similarity 99.9%; Pred. No. 2.7e-284;
Matches 1164; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ATGCCCGCGGCTCTCTGGGACCCGCGGTTACCTGCTCTACTGCGCCCTGCAATGGGCGC 60
Db 136 ATGCCCGCGGCTCTCTGGGACCCGCGGTTACCTGCTCTACTGCGCCCTGCAATGGGCGC 195
QY 61 TTTGGGAACCAAGCGCGATCATTCTTTGGGCTCTCTGGCAATTTGCAAGCTGCTAAACCGT 120
Db 196 TTTGGGAACCAAGCGCGATCATTCTTTGGGCTCTCTGGCAATTTGCAAGCTGCTAAACCGT 255
QY 121 ACCTTGGCTGCTCTCTGGGACCCGCGGTTACCTGCTCTACTGCGCCCTGCAATGGGCGC 180
Db 256 ACCTTGGCTGCTCTCTGGGACCCGCGGTTACCTGCTCTACTGCGCCCTGCAATGGGCGC 315
QY 181 CATGTGTCTTACCAGAGTACTTCAAGCTGGAGGCCCTCCAGGCTTACCATCGGCTCATC 240
Db 316 CATGTGTCTTACCAGAGTACTTCAAGCTGGAGGCCCTCCAGGCTTACCATCGGCTCATC 375
QY 241 AGCTTGGAGGATTTTCATGGAAGTGGCACTCCACCCACTGGGCCCTTGAGAGCGGGTG 300
Db 376 AGCTTGGAGGATTTTCATGGAAGTGGCACTCCACCCACTGGGCCCTTGAGAGCGGGTG 435
QY 301 GCATACCTCTTTGAGGTGGCAGCCAGCGAGCCAGATAGAGACCTGCCCCCATGAAG 360
Db 436 GCATACCTCTTTGAGGTGGCAGCCAGCGAGCCAGATAGAGACCTGCCCCCATGAAG 495
QY 361 GAAGGAAACCCCTTTGGGCCAATTCGGGATCAGTTTCAATGTGAGTTTCAACAAGTCGGAG 420
Db 496 GAAGGAAACCCCTTTGGGCCAATTCGGGATCAGTTTCAATGTGAGTTTCAACAAGTCGGAG 555
QY 421 CTTTATTACAGGATTTCTTTCATGCTTCTTCAAGAGAAATGAGGACGAGATTTTCT 480
Db 556 CTTTATTACAGGATTTCTTTCATGCTTCTTCAAGAGAAATGAGGACGAGATTTTCT 615
QY 481 CCAAGGAAACATCCGGTCTTCCCTGCCAGAGCCCGAGCCAGTTCCTCCGTCCTAGNA 540
Db 616 CCAAGGAAACATCCGGTCTTCCCTGCCAGAGCCCGAGCCAGTTCCTCCGTCCTAGAG 675
QY 541 GAACACAGGCCCACTACAGAAGTACATGTTGCTCAGACGAATGTTGAAGACGGGAGAG 600
Db 676 GAACACAGGCCCACTACAGAAGTACATGTTGCTCAGACGAATGTTGAAGACGGGAGAG 735
QY 601 GCCAGATTTCATGCCACCTTGTCCGGCCCTATGTGGGCATTTCATCTGCGCATTCGGCTCT 660
Db 736 GCCAGATTTCATGCCACCTTGTCCGGCCCTATGTGGGCATTTCATCTGCGCATTCGGCTCT 795
QY 661 GACTCGAAGACGCTGTGCCATGCTGAAGGACGGGACTGACAGGCTCGCACTTCATGGCC 720
Db 796 GACTCGAAGACGCTGTGCCATGCTGAAGGACGGGACTGACAGGCTCGCACTTCATGGCC 855
QY 721 TCTCCGCACTGTGTGGGCTACAGCCGACAGCAGCGCCGCCCTCACGATGACTATGTGC 780
```



JOURNAL Patent: JP 2001527389-A 9 25-DEC-2001;  
GENENTECH INC  
COMMENT OS Unidentified  
PN JP 2001527389-A/9  
PD 25-DEC-2001  
PF 17-DEC-1997 JP 1998532877  
PR 31-JAN-1997 US 08/792498,26-NOV-1997 US 08/978741 PI  
PC YANG WANG,MICHAEL W SPELTMAN  
C12N15/54,C12N9/10,C07K16/40  
CC Strandedness: Single;  
CC Topology: Linear;  
CC O-fucosyltransferase  
FH Key  
FT source Location/Qualifiers  
FT 1..1100  
FT /organism='Unidentified'.  
FEATURES Location/Qualifiers  
source 1..1100  
/organism='unidentified'  
/mol\_type='genomic DNA'  
/db\_xref='taxon:32644'

ORIGIN

Query Match 72.7%; Score 1100; DB 6; Length 1100;  
Best Local Similarity 100.0%; Pred. No. 3.4e-268;  
Matches 1100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCCGCGGGCTTCCTGGAGACCGCGCGGTAACTCTACTTGCCTCTCTGCCCCATGAGTTGCAAAACCGT 60  
DB 1 ATGCCGCGGGCTTCCTGGAGACCGCGCGGTAACTCTACTTGCCTCTCTGCCCCATGAGTTGCAAAACCGT 60

QY 61 TTGTGGAAACAGCGCCGATCATCTTTTGCGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120  
DB 61 TTGTGGAAACAGCGCCGATCATCTTTTGCGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120

QY 121 ACCTTGGCTGTCCTCTCTGGATTGATGACCAGCATCAAGACCTCTTTTCAACAACCTC 180  
DB 121 ACCTTGGCTGTCCTCTCTGGATTGATGACCAGCATCAAGACCTCTTTTCAACAACCTC 180

QY 181 CATGTGCTCTACAGAAGTACTTCAAGCTGGAGCCCTCCAGCTTACCATCGGTGATC 240  
DB 181 CATGTGCTCTACAGAAGTACTTCAAGCTGGAGCCCTCCAGCTTACCATCGGTGATC 240

QY 241 AGCTTGAGGATTTTCATGGAAGCTGGCACCCACCTGCGCCCTCTGAGAAGCGGTG 300  
DB 241 AGCTTGAGGATTTTCATGGAAGCTGGCACCCACCTGCGCCCTCTGAGAAGCGGTG 300

QY 301 GCATATCTGTTGAGTGGCGCCAGCGAAGCCAGATAAGAACAGTGCCTCCATGAAG 360  
DB 301 GCATATCTGTTGAGTGGCGCCAGCGAAGCCAGATAAGAACAGTGCCTCCATGAAG 360

QY 361 GAAGAAACCCCTTTGGCCCATCTTGGGATCAGTTTCACTGAGTTTCAACAAGTCGGAG 420  
DB 361 GAAGAAACCCCTTTGGCCCATCTTGGGATCAGTTTCACTGAGTTTCAACAAGTCGGAG 420

QY 421 CTTTTTACAGGATTTCTTCTAGTGTCTTCTACAGAGAACATGAGGACGAGATTTTCT 480  
DB 421 CTTTTTACAGGATTTCTTCTAGTGTCTTCTACAGAGAACATGAGGACGAGATTTTCT 480

QY 481 CCNAAGAACATCCGCTGCTTGGCCCTGCGAGAGCCAGCCAGTTCCTCGTCTTAGAA 540  
DB 481 CCNAAGAACATCCGCTGCTTGGCCCTGCGAGAGCCAGCCAGTTCCTCGTCTTAGAA 540

QY 541 GAACACAGCCACTACAGAAGTACATGTPATGCTCAGACGAATGCTCAAGACGGAGAG 600  
DB 541 GAACACAGCCACTACAGAAGTACATGTPATGCTCAGACGAATGCTCAAGACGGAGAG 600

QY 601 GCCCAGATTATGCCACCTTGTCCGGCCCTATGTTGGGCAATTCATCTGCGCAATGGCTCT 660  
DB 601 GCCCAGATTATGCCACCTTGTCCGGCCCTATGTTGGGCAATTCATCTGCGCAATGGCTCT 660

QY 661 GACTGAAGAAAGCTGTGCCATGCTGAAGACGGGATGCAAGGCTGCGCATTTATGGCC 720  
DB 661 GACTGAAGAAAGCTGTGCCATGCTGAAGACGGGATGCAAGGCTGCGCATTTATGGCC 720

Db 250 CATGTGCTCTACCAAGAGTACTTAAGCTGGAGCCCTCCAGGCTTTACCATGGGTCATC 309  
Qy 241 AGCTTGGAGGATTTTCATGAGAGAGTGGACCCACCCACCTGGCCCCCTGGAGAAGCGGGTG 300  
Db 310 AGCTTGGAGGATTTTCATGAGAGAGTGGACCCACCCACCTGGCCCCCTGGAGAAGCGGGTG 369  
Qy 301 GCATACCTGCTTTGAGGTGGCAGCCAGCAGGAGCCAGATAGAGAGACGTGCCCATGAAG 360  
Db 370 GCATACCTGCTTTGAGGTGGCAGCCAGCAGGAGCCAGATAGAGAGACGTGCCCATGAAG 429  
Qy 361 GAAGGAACCCCTTTCGCCCATCTTCGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420  
Db 430 GAAGGAACCCCTTTCGCCCATCTTCGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 489  
Qy 421 CTTTTTACAGGCAATTCCTTCAGTGCCTTCCTACAGAAACATAGAGAGCCAGAGATTTCT 480  
Db 490 CTTTTTACAGGCAATTCCTTCAGTGCCTTCCTACAGAAACATAGAGAGCCAGAGATTTCT 549  
Qy 481 CAAAGGAACATCCGGTGTTCCTTCAGGAGCCGCCAGCCAGTTCCTCGTCTTAGAA 540  
Db 550 CAAAGGAACATCCGGTGTTCCTTCAGGAGCCGCCAGCCAGTTCCTCGTCTTAGAG 609  
Qy 541 GAACACAGGCCACTACAGAAATGATGTATGTGTCAGACGAAATGTTGAAGACGGAGAG 600  
Db 610 GAACACAGGCCACTACAGAAATGATGTATGTGTCAGACGAAATGTTGAAGACGGAGAG 669  
Qy 601 GCCCAGATTCATGCCACCTTGTCCGCCCTATGTGGGCAATCATCTGGCAATTCATCTGGCTCT 660  
Db 670 GCCCAGATTCATGCCACCTTGTCCGCCCTATGTGGGCAATTCATCTGGCAATTCATCTGGCTCT 729  
Qy 661 GACTGGAGAACCCCTTGTCCCATCTGTAAGCAGCGGACTGCAGGCTCCGACTTCATGSGCC 720  
Db 730 GACTGGAGAACCCCTTGTCCCATCTGTAAGCAGCGGACTGCAGGCTCCGACTTCATGSGCC 789  
Qy 721 TCTCCGACGTGTGGGTGTACAGCGGAGCAGACGGCCGCCCTCAAGTACTATGTGC 780  
Db 790 TCTCCGACGTGTGGGTGTACAGCGGAGCAGACGGCCGCCCTCAAGTACTATGTGC 849  
Qy 781 CTGCTGACCTGAAGAGAGATCAGAGGGCTGTGAAGCTCTGGGTGAGGTGCGCTGATGCC 840  
Db 850 CTGCTGACCTGAAGAGAGATCAGAGGGCTGTGAAGCTCTGGGTGAGGTGCGCTGATGCC 909  
Qy 841 CAGTCGGTCTACGTGTCTAATTCAGAGAGTTATGTCTGAGTCTCAACAGCTCTTC 900  
Db 910 CAGTCGGTCTACGTGTCTAATTCAGAGAGTTATGTCTGAGTCTCAACAGCTCTTC 969  
Qy 901 AAAGGGAAGGTGAAGGTGGTGAAGCCTGAAGCCTGAGGTGGCCAGGTGAGCTGTATCATC 960  
Db 970 AAAGGGAAGGTGAAGGTGGTGAAGCCTGAAGCCTGAGGTGGCCAGGTGAGCTGTATCATC 1029  
Qy 961 CTCGCCCAAGCCAGCACCTTTATTTGGCAACTGTCTCTCTTCACTGCTTTGTGAAG 1020  
Db 1030 CTCGCCCAAGCCAGCACCTTTATTTGGCAACTGTCTCTCTTCACTGCTTTGTGAAG 1089  
Qy 1021 CGGGAGCGGACCTCCAGGGGAGGCGCTCTTCTTCTTCGATGACAGAGCCCTTAAG 1080  
Db 1090 CGGGAGCGGACCTCCAGGGGAGGCGCTCTTCTTCTTCTTCGATGACAGAGCCCTTAAG 1149  
Qy 1081 CTGCGGACGAGTTCTGA 1098  
Db 1150 CTGCGGACGAGTTCTGA 1167

RESULT 15  
AJ781500  
LOCUS  
DEFINITION Pan troglodytes mRNA for protein-O-fucosyltransferase 1 (fut12 gene).  
VERSION AJ781500  
KEYWORDS fut12 gene; pofut1 gene; protein-O-fucosyltransferase 1.  
SOURCE Pan troglodytes (chimpanzee)  
ORGANISM Pan troglodytes

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.  
1 Martinez-Duncker,I., Mollicone,R., Candellier,J.J., Breton,C. and Oriol,R.  
A new superfamily of protein-O-fucosyltransferases, alpha2-fucosyltransferases, and alpha6-fucosyltransferases: phylogeny and identification of conserved peptide motifs Glycobiology 13 (12), 1C-5C (2003)  
12966037  
2 Martinez-Duncker,I., Oriol,R. and Mollicone,R.  
Phylogeny of fucosyltransferases  
Unpublished  
3 (bases 1 to 1167)  
Oriol,R.  
Direct Submission  
Submitted (06-JUL-2004) Oriol R., US04, Inserm, 16 Av. Paul Vaillant-Couturier, 94807, FRANCE  
Location/Qualifiers  
1. 1167  
/organism="Pan troglodytes"  
/mol\_type="mRNA"  
/db\_xref="taxon:9598"  
1. 1167  
/gene="fut12"  
1. 1167  
/gene="fut12"  
/function="adds fucose to Ser or Thr"  
/note="pofut1"  
/codon\_start=1  
/product="protein-O-fucosyltransferase 1"  
/protein\_id="CAH03712.1"  
/db\_xref="GI:50057070"  
/translation="MGAAMARPLSVSFLILLPLCPMAGSDPAGVLLYCPWGRF QNQHDFSLAFKALLNRTLAVFPWIEYQHHPDPTNLHVSQYKXKLEPLQAYHRV ISLEDFMEKATLHWPPEKRVAYCFVAOORSDDKTKPMKSGNFFGFWQDFHVSN AKTEFTGISFASVIREQWSRFSPEKRPVLAALPAQFPVLEERKPLQKTNWSDEN VKTGEQIHAHLVRETYGHLRIGSDWKNACMLKDTAGSHFMAPOCVGTSRSTAA PLTMCLDPLKEIORAVKLWRSIDQSVYATDESIVPELQQLFKGKVVSLAP EVAQVDLYILGDADHPIGNVCSFTAFVKRERDLQGRPSFFGMDRPPKLDEF"

Query Match 71.8%; Score 1086.8; DB 9; Length 1167;  
Best Local Similarity 99.4%; Pred. No. 7.7e-265;  
Matches 1091; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 ATGCCCGCGGGCTCCTGGGACCGCGCGGTTTACCTGTCTACTGCCCCCTGCATGGGGCGC 60  
Db 70 ATGCCCGCGGGCTCCTGGGACCGCGCGGTTTACCTGTCTACTGCCCCCTGCATGGGGCGC 129  
Qy 61 TTTGGGAACAGGCGCATCATCTTTGGGCTCTCTGGCAITTTGCAAGCTGCTAAACCGT 120  
Db 130 TTTGGGAACAGGCGCATCATCTTTGGGCTCTCTGGCAITTTGCAAGCTGCTAAACCGT 189  
Qy 121 ACCTTGGCTGCTCCTCTTGGATTGAGTACCAGCATCACAGCTCTCTTTACCACCACTC 180  
Db 190 ACCTTGGCTGCTCCTCTTGGATTGAGTACCAGCATCACAGCTCTCTTTACCACCACTC 249  
Qy 181 CATGTGCTTACCAAGAGTACTTCAAGCTGGAGCGCTCCAGGCTTACCATTGGGTGATC 240  
Db 250 CATGTGCTTACCAAGAGTACTTCAAGCTGGAGCGCTCCAGGCTTACCATTGGGTGATC 309  
Qy 241 AGCTTGGAGGATTTTCATGGAGAGCTGGGACCCACCACTGGCCCCCTGAGAAGCGGGTG 300  
Db 310 AGCTTGGAGGATTTTCATGGAGAGCTGGGACCCACCACTGGCCCCCTGAGAAGCGGGTG 369  
Qy 301 GCATACCTGCTTTGAGGTGGCAGCCAGCAGGAGCCAGATAGAGAGACGTGCCCATGAAG 360  
Db 370 GCATACCTGCTTTGAGGTGGCAGCCAGCAGGAGCCAGATAGAGAGACGTGCCCATGAAG 429  
Qy 361 GAAGGAACCCCTTTCGCCCATCTTCGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420

```
Db 430 GAAGGAACCCCTTTGGCCCATTTCTGGGATCAGTTTCATGTAGTTTCAACAAGTCGGAG 489
Qy 421 CTTTTCACAGGCAATTCCTTCAGTCTTCTACAGAGAACATGGAGCCAGAGATTTTCT 480
Db 490 CTTTTCACAGGCAATTCCTTCAGTCTTCTACAGAGAACATGGAGCCAGAGATTTTCT 549
Qy 481 CCAAGGAACATCCGGTCTTGGCCCTGCGAGAGCCCGAGCCAGTTCCCGTCTAGAA 540
Db 550 CCAAGGAACATCCGGTCTTGGCCCTGCGAGAGCCCGGCCAGTTCCCGTCTCTGGAG 609
Qy 541 GAACACAGCCCACTACAGAGTACATGTTATGTTGAGAGAAATGGTGAAGACGGGAGAG 600
Db 610 GAACACAGCCCACTACAGAGTACATGTTATGTTGAGAGAAATGGTGAAGACGGGAGAG 669
Qy 601 GCCCAGATTCATGCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGATTGGCTCT 660
Db 670 GCCCAGATTCATGCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGATTGGCTCT 729
Qy 661 GACTGGAAGAACGCTGTGCCATGCTGAAGACGGGACTGACGGCTCGCACTTCATGGCC 720
Db 730 GACTGGAAGAACGCTGTGCCATGCTGAAGACGGGACCGCAGGCTCGCACTTCATGGCC 789
Qy 721 TCTCCGAGTGTGTGGGCTACAGCCGAGCAGCGGCCCTCAGCATGACTATGTGC 780
Db 790 TCTCCGAGTGTGTGGGCTACAGCCGAGCAGCGGCCCTCAGCATGACTATGTGC 849
Qy 781 CTGCTGACCTGAAGGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGGTCTGGATGCC 840
Db 850 CTGCTGACCTGAAGGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGGTCTGGATGCC 909
Qy 841 CAGTCGGTCTACGTGTACTGATTCCGAGAGTTATGTGCTGAGCTCCAAACAGCTCTTC 900
Db 910 CAGTCGGTCTATGTTGCTACTGATTCCGAGAGTTATGTGCTGAGCTCCAAACAGCTCTTC 969
Qy 901 AAAGGGAAGGTGAAGGTGCTGAGCCTGAAGCCTGAGGTGGCCCGAGGTGCACTGTATC 960
Db 970 AAAGGGAAGGTGAAGGTGCTGAGCCTGAAGCCTGAGGTGGCCCGAGGTGCACTGTATC 1029
Qy 961 CTCGCCCAAGCCGACCACTTTATTGGCAACTGTGTCTCTCTTCACTGCTTTGTGAAG 1020
Db 1030 CTCGCCCAAGCTGACCACTTTATTGGCAACTGTGTCTCTCTTCACTGCTTTGTGAAG 1089
Qy 1021 CGGAGCGGGACCTCCAGGGAGGCGCTCTTCTTTTCGGCATGGACAGGCCCTTAAG 1080
Db 1090 CGGAGCGGGACCTCCAGGGAGGCGCTCTTCTTTTCGGCATGGACAGGCCCTTAAG 1149
Qy 1081 CTGCGGACGAGTTCTGA 1098
Db 1150 CTGCGGACGAGTTCTGA 1167
```

Search completed: November 9, 2005, 07:59:57  
Job time : 6834.38 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2005, 23:23:13 ; Search time 883.842 Seconds  
(without alignments)  
10140.369 Million cell updates/sec

Title: US-09-774-954-1  
Perfect score: 1514  
Sequence: 1 ATGCCCGCGGCTCTCTGGCA.....TTTTGAAATCAATGAATTC 1514

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_16Dec04:\*  
1: geneseqn1980s:\*  
2: geneseqn1980s:\*  
3: geneseqn2000s:\*  
4: geneseqn2001as:\*  
5: geneseqn2001bs:\*  
6: geneseqn2002as:\*  
7: geneseqn2002bs:\*  
8: geneseqn2003as:\*  
9: geneseqn2003bs:\*  
10: geneseqn2003cs:\*  
11: geneseqn2003ds:\*  
12: geneseqn2004as:\*  
13: geneseqn2004bs:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1514	100.0	1514	2	AAV65632 Human hea
2	1512.4	99.9	11284	2	AAV65633 Plaemid c
3	1510.8	99.8	5218	13	ACN40746 Tumour-as
4	1509.8	99.7	5266	10	ADF81754 Leukaemia
5	1460.2	96.4	4230	4	AAK51510 Human pol
6	1455.4	96.1	4850	4	AAK52494 Human pol
7	1447.4	95.6	5009	2	AAV65634 First Eco
8	854.8	56.5	915	4	AAV65632 Human sec
9	608.6	40.2	10331	4	AAK79915 Human imm
10	608.6	40.2	10331	4	AAK79915 Human imm
11	442.8	29.2	477	9	ACH35532 Human rep
12	441	29.1	479	9	ACH35532 Human end
13	406.2	26.8	610	13	ADQ56824 Novel can
14	165.4	11.0	1209	4	ABL13947 Drosophil
15	150.4	9.9	3264	4	ABL13946 Drosophil
16	123.2	8.1	1545	4	AAF23894 Human sec
17	108.6	7.2	3793	4	ABL13894 Drosophil
18	71.8	4.7	3567	4	ABL13952 Drosophil
19	60	4.0	60	6	ABN44539 Human spl
20	56	3.7	760	3	AAZ98210 Human sig

21	50	3.3	2000	8	ADA71938	Rice gene
22	46.2	3.1	2000	8	ADA71938	Rice gene
23	42.4	2.8	1438	8	ACA31561	Prokaryot
24	41.2	2.7	114793	4	AAO08215	Human gen
25	40.6	2.7	558	5	AAH97782	Murine 7-
26	40.6	2.7	2036	13	ADT44716	Bacterial
27	40.4	2.7	5019	13	ADQ80232	Brain-spe
28	40.4	2.7	5704	12	ADO84111	Human tum
29	40.4	2.7	5932	10	ADB75488	Prostate
30	40	2.6	40	2	AAV65637	Primer am
31	40	2.6	40	2	AAV65635	Northern
32	40	2.6	1554	8	ADA69815	Rice gene
33	39	2.6	372	8	ACA43606	Prokaryot
34	39	2.6	1079	10	ADC37615	Human nuc
35	39	2.6	1131	11	ABD01472	Pseudomon
36	39	2.6	1209	3	AAA61286	Human sec
37	39	2.6	1209	8	ADA40219	Human sec
38	39	2.6	1209	10	ADD37708	Human sec
39	39	2.6	1209	10	ADA56378	Gene enco
40	39	2.6	1353	11	ABD01484	Pseudomon
41	39	2.6	1404	10	ADC16737	Human mod
42	39	2.6	1528	4	AAH15283	Human CDN
43	39	2.6	1555	4	AAAS4695	Human ful
44	39	2.6	1566	4	AAAS01567	Human sec
45	39	2.6	1575	6	ADI16595	Human NOV

ALIGNMENTS

RESULT 1  
AAV65632  
ID AAV65632 standard; DNA; 1514 BP.

AC AAV65632;

DT 16-DEC-1998 (first entry)

DE Human heart O-fucosyltransferase encoding DNA.

XX O-fucosyltransferase; epidermal growth factor; EGF; glycosylation;  
KW O-fucose; inhibitor; sensory neuron; retinal neuron; human; heart; ss.

OS Homo sapiens.

XX Key Location/Qualifiers

FT misc\_feature 1..1100

FT /tag= b

FT /note= "this actively expressed O-fucosyltransferase  
sequence is claimed for in claim 9"

FT CDS 1..1098

FT /tag= a

FT /product= "human heart O-fucosyltransferase"

XX WO9833924-A1.

PD 06-AUG-1998.

XX 17-DEC-1997; 97WO-US023401.

XX 31-JAN-1997; 97US-00792498.

XX 26-NOV-1997; 97US-00978741.

XX (GETH ) GENENTECH INC.

XX Wang Y, Spellman MW;

XX WPI; 1998-437477/37.

XX P-PSDB; AAW80571.

XX Human O-fucosyltransferase able to glycosylate epidermal growth factor  
domains - useful for diagnosis and treatment of diseases involving  
overexpression of the enzyme.







RESULT 3  
ACN40746  
ID ACN40746 standard; cDNA; 5218 BP.  
XX  
XX ACN40746;  
XX AC  
XX 18-NOV-2004 (first entry)  
XX  
XX Tumour-associated antigenic target (TAT) cDNA DNA326813, SEQ ID NO:5725.  
XX  
XX Tumour-associated antigenic target; TAT; human; overexpression; cancer;  
XX Tumour; diagnosis; cell proliferative disorder; breast cancer;  
XX colorectal cancer; lung cancer; ovarian cancer; liver cancer;  
XX central nervous system cancer; bladder cancer; pancreatic cancer;  
XX cervical cancer; melanoma; leukaemia; hybridisation probe;  
XX chromosome identification; chromosome mapping; gene mapping;  
XX gene therapy; cytosstatic; gene; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO2004030615-A2.  
XX  
XX 15-APR-2004.  
XX  
XX 29-SEP-2003; 2003WO-US028547.  
XX  
XX 02-OCT-2002; 2002US-0414971P.  
XX  
XX (GETH ) GENENTECH INC.  
XX  
XX Wu TD, Zhang Z, Zhou Y;  
XX  
XX WPI: 2004-347921/32.  
XX P-PSDB; ABM82223.  
XX  
XX New tumor-associated antigenic target polypeptides and nucleic acids,  
XX useful in preparing a medicament for treating or detecting a  
XX proliferative disorder, e.g. breast, lung, colorectal, ovarian or  
XX prostate cancer or tumor.  
XX  
XX Claim 1; SEQ ID NO 5725; 7273pp; English.  
XX  
XX The invention relates to human tumour-associated antigenic target (TAT)  
XX polypeptides, and their related nucleic acids. The TAT polypeptides are  
XX overexpressed in cancer tissues compared to normal tissues, and may thus  
XX serve as effective targets for the diagnosis and treatment of cancer in  
XX mammals. The invention also relates to nucleic acid and polypeptide  
XX sequences at least 80% identical to the TAT nucleic acids and  
XX polypeptides; expression vectors and host cells comprising a TAT nucleic  
XX acid; an antibody specific for a TAT polypeptide; a peptide or organic  
XX molecule which binds to a TAT polypeptide; fusion proteins comprising a  
XX TAT polypeptide; and methods and compositions for the treatment or  
XX diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,  
XX antibodies, antagonists, binding molecules and compositions are useful  
XX for diagnosing or treating a cell proliferative disorder associated with  
XX increased TAT expression, particularly cancers such as breast cancer,  
XX colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder  
XX cancer, pancreatic cancer, cervical cancer, cancers of the central  
XX nervous system, melanoma and leukaemia. TAT nucleic acids may further be  
XX used as hybridisation probes, in chromosome and gene mapping, in  
XX chromosome identification and in gene therapy. The present sequence  
XX represents a TAT nucleic acid of the invention  
XX  
XX Sequence 5218 BP; 1255 A; 1274 C; 1334 G; 1355 T; 0 U; 0 Other;  
XX  
XX Query Match 99.8%; Score 1510.8; DB 13; Length 5218;  
XX Best Local Similarity 99.9%; Pred. No. 0;  
XX Matches 1512; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
XX  
XX 1 ATGCCCGGGGCTCTGGGACCCGGGTTACCTGCTTACTGCTCTACTGCTGATGGGGCGC 60  
XX 119 ATGCTCGGGGCTCTGGGACCCGGGTTACCTGCTTACTGCTCTACTGCTGATGGGGCGC 178  
XX

QY 61 TTGGGAAACAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAGCTGTAAACCGT 120  
DB 179 TTGGGAAACAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAGCTGTAAACCGT 238  
QY 121 ACCTTGGCTGTCCCTCTCTTGGATTGAGTACAGCATCAAGGCTCTCTTTCACCAACCTC 180  
DB 239 ACCTTGGCTGTCCCTCTCTTGGATTGAGTACAGCATCAAGGCTCTCTTTCACCAACCTC 298  
QY 181 CATGTGCTCTACCAAGAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATGGGTCTATC 240  
DB 299 CATGTGCTCTACCAAGAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATGGGTCTATC 358  
QY 241 ACCTTGGAGGATTTTCATGGAGAGCTGSCACCCACCTGCGCCCTCAGAAAGCGGGTG 300  
DB 359 AGCTTGGAGGATTTTCATGGAGAGCTGSCACCCACCTGCGCCCTCAGAAAGCGGGTG 418  
QY 301 GCATCTCTCTTGGAGTGGAGCCGACGAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 360  
DB 419 GCATCTCTCTTGGAGTGGAGCCGACGAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 478  
QY 361 GAAGGAAACCCCTTTGGCCCATCTCTGGGATCATGTTTCAATGTAGTTTCAACAGTGGAG 420  
DB 479 GAAGGAAACCCCTTTGGCCCATCTCTGGGATCATGTTTCAATGTAGTTTCAACAGTGGAG 538  
QY 421 CTTTTCAGGCAATTTCTCTTCAAGTCTTCTTCAAGAAACAATGGAGCCAGAGATTTTCT 480  
DB 539 CTTTTCAGGCAATTTCTCTTCAAGTCTTCTTCAAGAAACAATGGAGCCAGAGATTTTCT 598  
QY 481 CCAAAGGAACATCCGGTGTCTTGGCCCTGCGAGAGCCCGCCAGCTTCCCGCTCTAGAA 540  
DB 599 CCAAAGGAACATCCGGTGTCTTGGCCCTGCGAGAGCCCGCCAGCTTCCCGCTCTAGAG 658  
QY 541 GAACACAGGCCACTACAGAAAGTACATGTTGTCAGACCAAAATGGTGAAGACGGGAGAG 600  
DB 659 GAACACAGGCCACTACAGAAAGTACATGTTGTCAGACCAAAATGGTGAAGACGGGAGAG 718  
QY 601 GCCGAGTATGATGCCACCTTGTCCGGCCCTATGTGGGCAATTCATCTGGGCAATTTGGTCT 660  
DB 719 GCCGAGTATGATGCCACCTTGTCCGGCCCTATGTGGGCAATTCATCTGGGCAATTTGGTCT 778  
QY 661 GACTGGAAGAACCGCTGTGCTGCTGTAAGAGCCGAGCTGCGACCTGCGACCTTCATGGCC 720  
DB 779 GACTGGAAGAACCGCTGTGCTGCTGTAAGAGCCGAGCTGCGACCTGCGACCTTCATGGCC 838  
QY 721 TCTCCGCAAGTGTGGGCTACAGCCGACAGCAGCGGCCCTCCAGATGACTATGTGC 780  
DB 839 TCTCCGCAAGTGTGGGCTACAGCCGACAGCAGCGGCCCTCCAGATGACTATGTGC 898  
QY 781 CTGCTGACCTGAAGAGGATCCAGAGCGGCTGTGAAGCTCTGGGTGAGGTGCTGTGATGCC 840  
DB 899 CTGCTGACCTGAAGAGGATCCAGAGCGGCTGTGAAGCTCTGGGTGAGGTGCTGTGATGCC 958  
QY 841 CAGTGGTCTACGTTGCTACTGATTCGAGAGTTATGTGCTGAGCTCCAAACAGCTCTTC 900  
DB 959 CAGTGGTCTACGTTGCTACTGATTCGAGAGTTATGTGCTGAGCTCCAAACAGCTCTTC 1018  
QY 901 AAAGGAAAGGTGAAGGTGGTGAAGCTGAGGCTGAGGTGGCCAGGTCGACCTGTACATC 960  
DB 1019 AAAGGAAAGGTGAAGGTGGTGAAGCTGAGGCTGAGGTCGAGGTGGCCAGGTCGATCATC 1078  
QY 961 CTCGGCCAGCGCCCACTTATTTGGCAACTGCTCTCTCTTCTACTGCTTGTGTGAAG 1020  
DB 1079 CTCGGCCAGCGCCCACTTATTTGGCAACTGCTCTCTCTTCTACTGCTTGTGTGAAG 1138  
QY 1021 CGGGAGCGGACCTCCAGGGGAGCGCTCTTCTTCTTCTGGCATGGACGGCCCTTAAAG 1080  
DB 1139 CGGGAGCGGACCTCCAGGGGAGCGCGCTCTTCTTCTTCTGGCATGGACGGCCCTTAAAG 1198  
QY 1081 CTGCGGAGCAGATTTCTGATTTCTGCGCGAGGACCAAGACCTCTCTATCTCTGGAGGACAG 1140  
DB 1199 CTGCGGAGCAGATTTCTGATTTCTGCGCGGAGGACCAAGACCTCTCTATCTCTGGAGGACAG 1258

QY 1141 AGTCTGAGCTGCTTCCAGCCAGGCTGGCAGCAGAGGTGCTCCGGGATGCAAACT 1200  
DB |||||  
1259 AGTCTGAGCTGCTTCCAGCCAGGCTGGCAGCAGAGGTGCTCCGGGATGCAAACT 1318  
QY 1201 CCTCTTCTCACCTGCCAAGATGGAGAGAGTGCAGAGGACCCCTCAAGAGGGAGACGC 1260  
DB |||||  
1319 CCTCTTCTCACCTGCCAAGATGGAGAGAGTGCAGAGGACCCCTCAAGAGGGAGACGC 1378  
QY 1261 TCCATATCCAGGGCAGATGAGTCTGAGGCTTCTAGGAGCAGGAGCATCTCCCATCGCAC 1320  
DB |||||  
1379 TCCATATCCAGGGCAGATGAGTCTGAGGCTTCTAGGAGCAGGAGCATCTCCCATCGCAC 1438  
QY 1321 GTGCTTCTGCTCTTCTGGGAATTTCTACATGGCAAGCAGTCCAGCTCCGCTTTCT 1380  
DB |||||  
1439 GTGCTTCTGCTCTTCTGGGAATTTCTACACACTGGCAAGCAGTCCAGCTCCGCTTTCT 1498  
QY 1381 GGTCCACTCTGCTCTGAGCAGCTGGGATGCTGAACTCTTCTAGAGAGATTTTATAGA 1440  
DB |||||  
1499 GGTCCACTCTGCTCTGAGCAGCTGGGATGCTGAACTCTTCTAGAGAGATTTTATAGA 1558  
QY 1441 GAGATTTCTATATTTTATACAAAGTCAATGACTATCTTAGAACTCTCTGTTTATAGA 1500  
DB |||||  
1559 GAGATTTCTATATTTTATACAAAGTCAATGACTATCTTAGAACTCTCTGTTTATAGA 1618  
QY 1501 AAATCATGAATTC 1514  
DB |||||  
1619 AAATCATGAATTC 1632

RESULT 4

ADFB1754  
ID ADFB1754 standard; DNA; 5266 BP.  
XX AC ADFB1754;  
XX DT 26-FEB-2004 (first entry)  
XX DE Leukaemia-related DNA sequence #2310.  
XX KW Cytostatic; Gene therapy; leukaemia; ss.  
XX OS Unidentified.  
XX PN WO2003039443-A2.  
XX PD 15-MAY-2003.  
XX PF 04-NOV-2002; 2002WO-EP012303.  
XX PR 05-NOV-2001; 2001EP-00126244.  
XX PR 30-APR-2002; 2002EP-00009758.  
XX XX (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.  
PA (UYLU-) UNIV LUDWIG MAXIMILIANS.  
PA (HAPE/) HAFERLACH T.  
PA (SCHO/) SCHOCH C.  
XX (KERN/) KERN W.  
XX  
PI Haferlach T, Schoch C, Kern W, Kohlmann A, Schnittger S, Dugas M;  
PI Eils R, Brors B, Mergenthaler S;  
XX  
XX WPI; 2003-505037/47.  
XX  
PT Determining the subtype of leukemia cells and whether a patient sample  
PT contains leukemia cells or other cells, useful for treating leukemia,  
PT comprises determining the expression profile of a group of markers in a  
PT patient sample.  
XX  
PS Disclosure; SEQ ID NO 2310; 2938pp; English.  
XX  
CC The present invention relates to a method (M1) for determining the  
CC subtype of leukaemia cells and whether a patient sample contains  
CC leukaemia cells. The method comprises determining the expression profile

CC of a group of markers in a patient sample. The method is useful for  
CC determining the presence of leukaemia cells, its types or subtypes, and  
CC for the preparation of a medicament for treating leukaemia.  
XX  
SQ Sequence 5266 BP; 1280 A; 1278 C; 1331 G; 1346 T; 0 U; 31 Other;  
Query Match 99.7%; Score 1509.8; DB 10; Length 5266;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1511; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 ATGCGCGGGGCTCTCGGAGCCCGGTTACTGCTCTACTGCGCCCTGCAATGGGGCGC 60  
DB |||||  
132 ATGCGTGGGGCTCTCGGAGCCCGGCTTACTGCTCTACTGCGCCCTGCAATGGGGCGC 191  
QY 61 TTTGGGAACCAAGCCGATCACCTTCTTGGGCTCTCTGGCATTTGCAAGCTGCTAAACCGT 120  
DB |||||  
192 TTTGGGAACCAAGCCGATCACCTTCTTGGGCTCTCTGGCATTTGCAAGCTGCTAAACCGT 251  
QY 121 ACCTTGGCTGCTCCCTCTTGGATTGAGTACAGCATCAAGCCCTCTTTCAACCAACCTC 180  
DB |||||  
252 ACCTTGGCTGCTCCCTCTTGGATTGAGTACAGCATCAAGCCCTCTTTCAACCAACCTC 311  
QY 181 CATGCTCTTACAGAACTACTTCAAGCTGGAGCCCTCCAGGCTTACATCGGGTCATC 240  
DB |||||  
312 CATGCTCTTACAGAACTACTTCAAGCTGGAGCCCTCCAGGCTTACATCGGGTCATC 371  
QY 241 AGCTTGGAGGATTTTCATGGAGAGCTGCACCCACCTGCGCCCTGAGAGCGGGTG 300  
DB |||||  
372 AGCTTGGAGGATTTTCATGGAGAGCTGCACCCACCTGCGCCCTGAGAGCGGGTG 431  
QY 301 GCATACTCTTGGAGTGGCAGCCAGCAAGCCAGATTAAGAACAGCTGCCCATGAAG 360  
DB |||||  
432 GCATACTCTTGGAGTGGCAGCCAGCAAGCCAGATTAAGAACAGCTGCCCATGAAG 491  
QY 361 GAAGAAACCCCTTTGGCCCATTTCTGGGATCAGTTTCAATGAGTTTCAACAAAGTCGAG 420  
DB |||||  
492 GAAGAAACCCCTTTGGCCCATTTCTGGGATCAGTTTCAATGAGTTTCAACAAAGTCGAG 551  
QY 421 CTTTTCAGGCAATTTCTTCTCAGTCTTCTTACAGAGAACATGAGGACGAGATTTCT 480  
DB |||||  
552 CTTTTCAGGCAATTTCTTCTCAGTCTTCTTACAGAGAACATGAGGACGAGATTTCT 611  
QY 481 CCAAGGAACATCCGGTCTTCTGCGCCCTGCGAGAGCCCGCCAGCTTCCCGCTCTAGAA 540  
DB |||||  
612 CCAAGGAACATCCGGTCTTCTGCGCCCTGCGAGAGCCCGCCAGCTTCCCGCTCTAGAG 671  
QY 541 GAACACAGGCCACTACAGAACTACATGTTTGGTTCAGACGAAATGGTGAAGACGGGAGAG 600  
DB |||||  
672 GAACACAGGCCACTACAGAACTACATGTTTGGTTCAGACGAAATGGTGAAGACGGGAGAG 731  
QY 601 GCCCAGATTTCATGCCACCTTGTCCGGCCCTATGTGGGCATTCATCTCGGCATTTGGCTCT 660  
DB |||||  
732 GCCCAGATTTCATGCCACCTTGTCCGGCCCTATGTGGGCATTCATCTCGGCATTTGGCTCT 791  
QY 661 GACTTGAAGAAGCGCTGTGCCATCTGAAGACCGGACTGCGAGGCTCCGCTTATGCCC 720  
DB |||||  
792 GACTTGAAGAAGCGCTGTGCCATCTGAAGACCGGACTGCGAGGCTCCGCTTATGCCC 851  
QY 721 TCTCCGAGTGTGTGGGTACAGCCGACGACAGCGCCCGCCCTCACGATGATATGTGC 780  
DB |||||  
852 TCTCCGAGTGTGTGGGTACAGCCGACGACAGCGCCCGCCCTCACGATGATATGTGC 911  
QY 781 CTGCTGACCTGAAGGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGGTGCGTGGATGCC 840  
DB |||||  
912 CTGCTGACCTGAAGGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGGTGCGTGGATGCC 971  
QY 841 CAGTGGGTCTACGTTTGTACTGATTTCCGAGAGTTATGTGCGCTGAGCTCCACAGCTCTTC 900  
DB |||||  
972 CAGTGGGTCTACGTTTGTACTGATTTCCGAGAGTTATGTGCGCTGAGCTCCACAGCTCTTC 1031  
QY 901 AAAGGAAGGTGAAGGTGGTGGAGCTGAAGCTGAGGTGGCGCCCGAGCTGACATC 960  
DB |||||  
1032 AAAGGAAGGTGAAGGTGGTGGAGCTGAAGCTGAGGTGGCGCCCGAGCTGACATC 1091





Db 446 AGAAGCGGGTGGCATATCTGCTTTAGGTGGCAGCCAGCGAAGCCGAGATAGGAAGCGT 505  
Qy 350 GCCCATGAAGGAAGAAACCCCTTTGGCCCATTTCTGGGATCAGTTTTCATGTGAGTTTCA 409  
Db 506 GCCCATGAAGGAAGAAACCCCTTTGGCCCATTTCTGGGATCAGTTTTCATGTGAGTTTCA 565  
Qy 410 ACAAGTGGAGCTTTTACAGCATTTCTCTTCAGTGTCTCTACAGAGAAACAATGGAGCC 469  
Db 566 ACAAGTGGAGCTTTTACAGCATTTCTCTTCAGTGTCTCTACAGAGAAACAATGGAGCC 625  
Qy 470 AGAGATTTTCTCCAGAGACATCCGCTGCTTCCCTGCGCAGGAGCCCGCAGCCAGTTCC 529  
Db 626 AGAGATTTTCTCCAGAGACATCCGCTGCTTCCCTGCGCAGGAGCCCGCAGCCAGTTCC 685  
Qy 530 CGCTCTAGAAAGAACACAGGCCCATACAGAAAGTACATGTTGTCAGACGAAATGGTGA 589  
Db 686 CGCTCTAGAAAGAACACAGGCCCATACAGAAAGTACATGTTGTCAGACGAAATGGTGA 745  
Qy 590 AGACGGGAGAGCCACAGATTATGATGCCCATCTGTCGCGCCCTATGTGGCAATTCATCTGC 649  
Db 746 AGACGGGAGAGCCACAGATTATGATGCCCATCTGTCGCGCCCTATGTGGCAATTCATCTGC 805  
Qy 650 GCATTGGCTCTGACTGGAAGAACCGCTGTGCATGCTGAAGACGGGACTGCAGGCTGCG 709  
Db 806 GCATTGGCTCTGACTGGAAGAACCGCTGTGCATGCTGAAGACGGGACTGCAGGCTGCG 865  
Qy 710 ACTTCATGGCTCTCCGCACTGTGGGCTACAGCCGAGCAGAGCGGCCCCCTCAGCA 769  
Db 866 ACTTCATGGCTCTCCGCACTGTGGGCTACAGCCGAGCAGAGCGGCCCCCTCAGCA 925  
Qy 770 TGACATATGCTGCTGCTGACCTGACCTGAGAGATCCAGAGGCTGTGAAAGCTCTGGGTGAGGT 829  
Db 926 TGACATATGCTGCTGCTGACCTGAGAGAGATCCAGAGGCTGTGAAAGCTCTGGGTGAGGT 985  
Qy 830 CGCTGGATCCCGAGTCGGTCTAGCTGTGCTACTGATCCGAGAGTATGTCCTGAGGCTCC 889  
Db 986 CGCTGGATCCCGAGTCGGTCTAGCTGTGCTACTGATCCGAGAGTATGTCCTGAGGCTCC 1045  
Qy 890 AACAGCTTTCAAGAGGAAGGTGAAGGTGTGAGCTGTGAGCTGTGAGCTGTGAGCTGTG 949  
Db 1046 AACAGCTTTCAAGAGGAAGGTGAAGGTGTGAGCTGTGAGCTGTGAGCTGTGAGCTGTG 1105  
Qy 950 ACTGTATCATCTCTCGGCCAAGCCGACACTTTATTGGCACTGTGTCTCTCTTCACTG 1009  
Db 1106 ACTGTATCATCTCTCGGCCAAGCCGACACTTTATTGGCACTGTGTCTCTCTTCACTG 1165  
Qy 1010 CTTTGTGAAGCGGAGCGGGAACCTCAGGGAGGCGCTCTTCTTTCTGGCATGGACA 1069  
Db 1166 CTTTGTGAAGCGGAGCGGGAACCTCAGGGAGGCGCTCTTCTTTCTGGCATGGACA 1225  
Qy 1070 GGCCCTTAAGCTCGGACAGGTTCGTATTTCTGGCCGAGCACAGACCTCTGATCTCT 1129  
Db 1226 GGCCCTTAAGCTCGGACAGGTTCGTATTTCTGGCCGAGCACAGACCTCTGATCTCT 1285  
Qy 1130 GGAGGACACAGACTGAGCTGGTCTCTCCAGCAGGCTGCGCAGCCAGAGGTGCTCCGG 1189  
Db 1286 GGAGGACACAGACTGAGCTGGTCTCTCCAGCAGGCTGCGCAGCCAGAGGTGCTCCGG 1345  
Qy 1190 GATTGAAACTCTCTTCTACCTGCGCAAGATGAGAGAGTCCAGGAGCCCTCAAG 1249  
Db 1346 GATTGAAACTCTCTTCTACCTGCGCAAGATGAGAGAGTCCAGGAGCCCTCAAG 1405  
Qy 1250 GAGGAGACGCTCCATATPCCAGGGGANTAGGACTTGCAGGTTCTTAGAGCAGAGCATC 1309  
Db 1406 GAGGAGACGCTCCATATPCCAGGGGANTAGGACTTGCAGGTTCTTAGAGCAGAGCATC 1465  
Qy 1310 TCCCATCGCAGCTGTCTTCTGCTGCTTCTGCTGCTTCTGCTGCTGCTGCTGCTGCTGCT 1369  
Db 1466 TCCCATCGCAGCTGTCTTCTGCTGCTTCTGCTGCTTCTGCTGCTGCTGCTGCTGCTGCT 1525  
Qy 1370 CTCCGTCTTCTGGTCCACTCTGCTCTGAGCAGCTGGGATGCTGAATCTTCTCAGAGAT 1429

Db 1526 CTCCTCTTCTGGTGCACCTCTGCTCTGAGCAGCCTGGATGCTGAACCTTTCAGAGAGAT 1585  
Qy 1430 TTTTATATAGAGAGATTTCTTAATAATTTTATGATACAAGGTCAATGATCTTCTAGAACTCTCT 1489  
Db 1586 TTTTATATAGAGAGATTTCTTAATAATTTTATGATACAAGGTCAATGATCTTCTAGAACTCTCT 1645  
Qy 1490 GTGGTTTTTGAAAATCATTTGAATTC 1514  
Db 1646 GTGGTTTTTGAAAATCATTTGAATTC 1670  
RESULT 7  
AAV65634  
ID AAV65634 standard; cDNA; 5009 BP.  
XX  
AC AAV65634;  
XX  
XX 16-DEC-1998 (first entry)  
XX First EcoRI nucleotide fragment of human KIAA0180.  
DE O-fucosyltransferase; epidermal growth factor; EGF; glycosylation;  
KW O-fucose; inhibitor; sensory neuron; retinal neuron; human; heart; ss.  
XX Homo sapiens.  
XX OS  
XX PN WO9833924-A1.  
XX PD 06-AUG-1998.  
XX PF 17-DEC-1997; 97WO-US023401.  
XX PR 31-JAN-1997; 97US-00792498.  
XX PR 26-NOV-1997; 97US-00978741.  
XX PA (GETH ) GENENTECH INC.  
XX PI Wang Y, Spellman MW;  
XX DR WPI: 1998-437477/37.  
XX DR P-PSDB; AAN80577.  
PT Human O-fucosyltransferase able to glycosylate epidermal growth factor  
PT domains - useful for diagnosis and treatment of diseases involving  
PT overexpression of the enzyme.  
PS Example 1; Fig 11; 90pp; English.  
XX This represents a first EcoRI nucleotide fragment of human KIAA0180. This  
CC 5009 basepairs partial cDNA encodes for a protein of unknown function  
CC from myelast cel line KG-1. The invention provides a human heart O-  
CC fucosyltransferase that can glycosylate an epidermal growth factor (EGF)  
CC domain of a polypeptide with an activated O-fucose residue. Inhibitors of  
CC O-fucosyltransferase, e.g. mutants with increased affinity for the EGF  
CC domains, are used in diagnosis and treatment of conditions associated  
CC with overexpression of O-fucosyltransferase, to promote survival of  
CC sensory (retinal) neurons. Probes based on EGF domain polypeptide are  
CC used to detect gene amplification and expression. The expression can also  
CC be determined at the protein level using antibodies specific for O-  
CC fucosyltransferase  
XX  
SQ Sequence 5009 BP; 1235 A; 1195 C; 1268 G; 1311 T; 0 U; 0 Other;

Query Match 95.6%; Score 1447.4; DB 2; Length 5009;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1446; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 66 GAACACAGGCGGATCATCTTCTGGGCTCTGGGATTTGCAAGCTGTAAACCGTACCTT 125  
Db 1 GAACACAGGCGGATCATCTTCTGGGCTCTGGGATTTGCAAGCTGTAAACCGTACCTT 60  
Qy 126 GGCTGTCCCTCTCTGGATGATGATACAGATCAAGCCTCTTTCACCACTCCATGT 185

Db 61 GCGTGTCCCTCCTTGAGTTAGTACACAGATCACAAGCCTCCTTTTACCACCACTCCATGT 120  
Qy 186 GTCTACCAAGAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGTCATCAGCTT 245  
Db 121 GTCTACCAAGAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGTCATCAGCTT 180  
Qy 246 GGAGGATTTATGGAGAGCTGGCAACCAACCCACTGGCCCCCTGAGAGCGGGTGGCATA 305  
Db 181 GGAGGATTTATGGAGAGCTGGCAACCAACCCACTGGCCCCCTGAGAGCGGGTGGCATA 240  
Qy 306 CTGCTTTGAGGTGGCAGCCGAGAGCCAGATGAAGAGCTGCCCATGAAGGAAG 365  
Db 241 CTGCTTTGAGGTGGCAGCCGAGAGCCAGATGAAGAGCTGCCCATGAAGGAAG 300  
Qy 366 AAACCCCTTTGGCCCATTTCTGGATCAGTTTCATGTGATTTCAACAAGTCGAGCTTTT 425  
Db 301 AAACCCCTTTGGCCCATTTCTGGATCAGTTTCATGTGATTTCAACAAGTCGAGCTTTT 360  
Qy 426 TACAGGCATTTCTCTTCACTGCTTCTTACAGAGAACATGGAGCCAGAGATTTTCTCCAAA 485  
Db 361 TACAGGCATTTCTCTTCACTGCTTCTTACAGAGAACATGGAGCCAGAGATTTTCTCCAAA 420  
Qy 486 GGACATCCGCTGCTTGGCCCTGCGAGAGCCCGCCAGCTTCCCGCTCTAGAGAAC 545  
Db 421 GGACATCCGCTGCTTGGCCCTGCGAGAGCCCGCCAGCTTCCCGCTCTAGAGAAC 480  
Qy 546 CAGGCACTACAGAGTACATGTTATGTTGATGTCAGACGAAATGTTGAAGCGGAGAGGCCCA 605  
Db 481 CAGGCACTACAGAGTACATGTTATGTTGATGTCAGACGAAATGTTGAAGCGGAGAGGCCCA 540  
Qy 606 GATTATGCCCATTTGTCGGCCCTATGTTGGGCAATTCATTCGGCATTTGGCTCTGACTG 665  
Db 541 GATTATGCCCATTTGTCGGCCCTATGTTGGGCAATTCATTCGGCATTTGGCTCTGACTG 600  
Qy 666 GAAGAACCCCTGTCGCTGTAAGGACGGAAGTCTGAGCTCGACTTCATTCAGGCTCTCC 725  
Db 601 GAAGAACCCCTGTCGCTGTAAGGACGGAAGTCTGAGCTCGACTTCATTCAGGCTCTCC 660  
Qy 726 GCAGTGTGTGGCTACAGCCGAGCAGACAGCGGCCCTCTCAGATGACTATGTGCTGCC 785  
Db 661 GCAGTGTGTGGCTACAGCCGAGCAGACAGCGGCCCTCTCAGATGACTATGTGCTGCC 720  
Qy 786 TGACCTGAAGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCTGCTGGATGCCAGTC 845  
Db 721 TGACCTGAAGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCTGCTGGATGCCAGTC 780  
Qy 846 GGTCTAGCTGTCTACTGATTCGAGAGTATGTCCTGAGCTCCAAAGCTCTTCAAGG 905  
Db 781 GGTCTAGCTGTCTACTGATTCGAGAGTATGTCCTGAGCTCCAAAGCTCTTCAAGG 840  
Qy 906 GAAGTGAAGTGTGAGCTGAGCCTGAAGCCTGAGTGGCCAGCTCGACCTGTATACCTCCGG 965  
Db 841 GAAGTGAAGTGTGAGCCTGAAGCCTGAGTGGCCAGCTCGACCTGTATACCTCCGG 900  
Qy 966 CCAAGCCGACCACTTTATTTGGCACTGTGTCTCTCTCTTCACTGCTTTGTGAAGCGGGA 1025  
Db 901 CCAAGCCGACCACTTTATTTGGCACTGTGTCTCTCTCTTCACTGCTTTGTGAAGCGGGA 960  
Qy 1026 CCGGACCTCAGGGAGCGCTCTTTCTTCTTGGGATGACAGGCCCCCTAAGCTGCG 1085  
Db 961 CCGGACCTCAGGGAGCGCTCTTTCTTCTTGGGATGACAGGCCCCCTAAGCTGCG 1020  
Qy 1086 GGACGAGTTCTGATTTCTGGCCGAGGACCAAGACCTCTGATCTGGAGGACCAAGTCT 1145  
Db 1021 GGACGAGTTCTGATTTCTGGCCGAGGACCAAGACCTCTGATCTGGAGGACCAAGTCT 1080  
Qy 1146 GAGCTGTCTCTTCCAGCCAGGCTGGCAGCCAGAGGTGCTCCGGGATTTGCAAACTCTCT 1205  
Db 1081 GAGCTGTCTCTTCCAGCCAGGCTGGCAGCCAGAGGTGCTCCGGGATTTGCAAACTCTCT 1140  
Qy 1206 TCTCACCTGCAAGATGAGAGAGTGTCCAGGAGCCCTCAAGGAGGAGAGCTCCAT 1265  
Db 1141 TCTCACCTGCAAGATGAGAGAGTGTCCAGGAGCCCTCAAGGAGGAGAGCTCCAT 1200

## RESULT 8

AAD07629

ID AAD07629 standard; cDNA; 915 BP.

XX AAD07629;

XX AC AAD07629;

DT 10-AUG-2001 (first entry)

XX DE Human secreted protein-encoding gene 18 cDNA clone HMWEY26, SEQ ID NO:69.

XX KW Human; secreted protein; proliferative disorder; cancer; tumour;

KW KW foetal abnormality; developmental abnormality; haematopoietic disorder;

KW KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;

KW KW inflammation; allergy; neurological disorder; Alzheimer's disease;

KW KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;

KW KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;

KW KW cardiovascular disorder; angiogenic disorder; kidney disorder;

KW KW gastrointestinal disorder; pregnancy-related disorder; gene therapy;

KW KW endocrine disorder; infection; wound healing; vulnery; cell culture;

KW KW chemotaxis; food additive; binding partner identification; chromosome 3;

XX KW ss.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

FT CDS 1..498

FT FT /\*tag= a

FT FT /product= "Human secreted protein precursor"

FT FT /transl\_except= (pos:4..6, aa:Xaa)

FT FT /transl\_except= (pos:10..12, aa:Xaa)

FT FT /transl\_except= (pos:13..15, aa:Xaa)

FT FT /transl\_except= (pos:442..444, aa:Xaa)

FT FT /note= "Xaa equals any of the naturally occurring L-amino acids; CDS does not include start codon"

FT FT /partial

FT FT 1..3

FT FT /\*tag= b

FT FT 4..495

FT FT /\*tag= c

FT FT /product= "Mature human secreted protein"

XX XX

PN WO200132676-A1.

XX PD 10-MAY-2001.

XX PF 25-OCT-2000; 2000WO-US029365.

XX XX 29-OCT-1999; 99US-0162237P.

PR 21-JUL-2000; 2000US-0219666P.

XX XX (HUMA-) HUMAN GENOME SCI INC.







PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225269P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226688P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 27-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235835P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236328P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-483426/52.  
XX  
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
XX useful for preventing, diagnosing and/or treating cancers and metastasis.  
XX  
XX Disclosure; SEQ ID NO 34727; 3071pp + Sequence Listing; English.  
XX  
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)  
XX amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic  
XX activity, and can be used in gene therapy and vaccine production. (I)  
XX proteins and polynucleotides may be used in the prevention, diagnosis and  
XX treatment of diseases associated with inappropriate (I) expression. For  
XX example, they may be used to treat disorders associated with decreased  
XX expression by rectifying mutations or deletions in a patient's genome  
XX that affect the activity of (I) by expressing inactive proteins or to  
XX supplement the patient's own production of (I). Additionally, (I)  
XX polynucleotides may be used to produce the secreted (I), by inserting the  
XX nucleic acids into a host cell and culturing the cell to express the  
XX protein. (I) proteins and polynucleotides may be used to prevent,  
XX diagnose and treat immune/hematopoietic-related diseases, especially

CC	cancers and cancer metastases of haematopoietic-derived cells. AAK64703	
CC	to AAK87694 represent human immune/haematopoietic antigen genomic	
CC	sequences from the present invention. AAK54942 to AAK54950 and AAK82169	
CC	represent sequences used in the exemplification of the present invention	
XX		
SQ	Sequence 10331 BP; 2718 A; 2405 C; 2599 G; 2609 T; 0 U; 0 Other;	
	Query Match 40.2%; Score 608.6; DB 4; Length 10331;	
	Best Local Similarity 98.6%; Pred. No. 4e-166;	
	Matches 614; Conservative 0; Mismatches 9; Indels 0; Gaps 0;	
Qy	892 CAGCTCTTCAAGGGAAGGTGAAGGTGGTGAAGCTCTGAAGCTCAGGTGGCCAGGTCCGAC 951	PD 02-AUG-2001.
Db	6127 CTGTCTTCTTCTGCTGAGGTGAAGGTGGTGAAGCTCTGAAGCTCAGGTGGCCAGGTCCGAC 6186	XX 17-JAN-2001; 2001WO-US001339.
Qy	952 CTGTATCTCTCGCCCAAGCCGACACTTTATTGGCACTGTCTCTCTCTCTCTCTCTCTCT 1011	PR 31-JAN-2000; 2000US-0179065P.
Db	6187 CTGTATCTCTCGCCCAAGCCGACACTTTATTGGCACTGTCTCTCTCTCTCTCTCTCTCT 6246	PR 04-FEB-2000; 2000US-0180628P.
Qy	1012 TTTGTGAAGCGGAGCGGAGCCTTCAGGGAGGCGCTCTCTTTCTTGGCGATGACACAGG 1071	PR 24-FEB-2000; 2000US-0184664P.
Db	6247 TTTGTGAAGCGGAGCGGAGCCTTCAGGGAGGCGCTCTCTTTCTTGGCGATGACACAGG 6306	PR 02-MAR-2000; 2000US-0186350P.
Qy	1072 CCCCTAAGCTGCGGAGCGAGTTCTGATTCTGGCCGAGCACAGACCCCTCTGATCCTCG 1131	PR 16-MAR-2000; 2000US-0189874P.
Db	6307 CCCCTAAGCTGCGGAGCGAGTTCTGATTCTGGCCGAGCACAGACCCCTCTGATCCTCG 6366	PR 17-MAR-2000; 2000US-0190076P.
Qy	1132 AGGGACACAGAGTCTGAGCTGGTCTCTCCAGCCAGGCTTGGCAGCCAGAGGTGCTCCGGGA 1191	PR 18-APR-2000; 2000US-0198123P.
Db	6367 AGGGACACAGAGTCTGAGCTGGTCTCTCCAGCCAGGCTTGGCAGCCAGAGGTGCTCCGGGA 6426	PR 19-MAY-2000; 2000US-0205515P.
Qy	1192 TTGCAAACTCTCTCTCTCACCTGCGCAAGATGAGAGTGCCAGGACCCCTCAAGGA 1251	PR 07-JUN-2000; 2000US-0209467P.
Db	6427 TTGCAAACTCTCTCTCACCTGCGCAAGATGAGAGTGCCAGGACCCCTCAAGGA 6486	PR 28-JUN-2000; 2000US-0214886P.
Qy	1252 GGGAGACCTCCATATCCAGGGCATAGGACTTGCAGGTTCTTAGGAGCAGGAGCATCTC 1311	PR 30-JUN-2000; 2000US-0215135P.
Db	6487 GGGAGACCTCCATATCCAGGGCATAGGACTTGCAGGTTCTTAGGAGCAGGAGCATCTC 6546	PR 07-JUL-2000; 2000US-0216647P.
Qy	1312 CCATCGCACGTCCTTCTGCTCTCTCTGGGAATTTCTCACACTGGCAAGCAGTCAGGCT 1371	PR 11-JUL-2000; 2000US-0216880P.
Db	6547 CCATCGCACGTCCTTCTGCTCTCTCTGGGAATTTCTCACACTGGCAAGCAGTCAGGCT 6606	PR 14-JUL-2000; 2000US-0217496P.
Qy	1372 CCGTCTCTGCTCCACTCTGCTCTGAGCAGCTGGGATGCTGAACCTTTCAGAGAGATTT 1431	PR 11-JUL-2000; 2000US-0218290P.
Db	6607 CCGTCTCTGCTCCACTCTGCTCTGAGCAGCTGGGATGCTGAACCTTTCAGAGAGATTT 6666	PR 26-JUL-2000; 2000US-0220963P.
Qy	1432 TTTTATAGAGAGATTTCTATATTTTATAGACAGTCTATGACTATCTTAGAATCTCTGT 1491	PR 26-JUL-2000; 2000US-0220964P.
Db	6667 TTTTATAGAGAGATTTCTATATTTTATAGACAGTCTATGACTATCTTAGAATCTCTGT 6726	PR 14-AUG-2000; 2000US-0224518P.
Qy	1492 GGTTTTCAAAATCATTGAATTC 1514	PR 14-AUG-2000; 2000US-0224519P.
Db	6727 GGTTTTCAAAATCATTGAATTC 6749	PR 14-AUG-2000; 2000US-0225213P.
		PR 14-AUG-2000; 2000US-0225214P.
		PR 14-AUG-2000; 2000US-0225266P.
		PR 14-AUG-2000; 2000US-0225267P.
		PR 14-AUG-2000; 2000US-0225268P.
		PR 14-AUG-2000; 2000US-0225270P.
		PR 14-AUG-2000; 2000US-0225447P.
		PR 14-AUG-2000; 2000US-0225757P.
		PR 14-AUG-2000; 2000US-0225758P.
		PR 14-AUG-2000; 2000US-0225759P.
		PR 18-AUG-2000; 2000US-0226279P.
		PR 22-AUG-2000; 2000US-0226681P.
		PR 22-AUG-2000; 2000US-0226688P.
		PR 22-AUG-2000; 2000US-0227182P.
		PR 23-AUG-2000; 2000US-0227009P.
		PR 30-AUG-2000; 2000US-0228924P.
		PR 01-SEP-2000; 2000US-0229287P.
		PR 01-SEP-2000; 2000US-0229343P.
		PR 01-SEP-2000; 2000US-0229344P.
		PR 01-SEP-2000; 2000US-0229345P.
		PR 05-SEP-2000; 2000US-0229509P.
		PR 05-SEP-2000; 2000US-0229513P.
		PR 06-SEP-2000; 2000US-0230437P.
		PR 06-SEP-2000; 2000US-0230438P.
		PR 08-SEP-2000; 2000US-0231242P.
		PR 08-SEP-2000; 2000US-0231243P.
		PR 08-SEP-2000; 2000US-0231244P.
		PR 08-SEP-2000; 2000US-0231413P.
		PR 08-SEP-2000; 2000US-0231414P.
		PR 08-SEP-2000; 2000US-0232080P.
		PR 08-SEP-2000; 2000US-0232081P.
		PR 12-SEP-2000; 2000US-0231968P.
		PR 14-SEP-2000; 2000US-0232397P.
		PR 14-SEP-2000; 2000US-0232398P.
		PR 14-SEP-2000; 2000US-0232399P.
		PR 14-SEP-2000; 2000US-0232400P.
		PR 14-SEP-2000; 2000US-0232401P.
		PR 14-SEP-2000; 2000US-0233063P.
		PR 14-SEP-2000; 2000US-0233064P.
		PR 14-SEP-2000; 2000US-0233065P.
		PR 21-SEP-2000; 2000US-0233066P.
		PR 21-SEP-2000; 2000US-0234274P.
		PR 25-SEP-2000; 2000US-0234997P.
		PR 25-SEP-2000; 2000US-0234998P.
		PR 26-SEP-2000; 2000US-0235484P.
		PR 27-SEP-2000; 2000US-0235834P.
		PR 27-SEP-2000; 2000US-0235836P.
		PR 29-SEP-2000; 2000US-0236327P.
		PR 29-SEP-2000; 2000US-0236367P.

RESULT 10  
AAL03344  
ID AAL03344 standard; DNA; 10331 BP.  
XX AAL03344;  
AC AAL03344;  
DT 21-NOV-2001 (first entry)  
XX Human reproductive system related antigen DNA SEQ ID NO: 6032.  
DE Human reproductive system related antigen; reproductive system disorder;  
XX Human; reproductive system related antigen; cancer; gene therapy; ds.  
XX cancer; gene therapy; ds.  
XX Homo sapiens.  
XX Homo sapiens.  
XX WO200155320-A2.  
XX

PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241828P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249219P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX  
PI Rosen CA, Baraah SC, Ruben SM;  
XX  
XX WPI; 2001-465570/50.  
DR

XX Isolated nucleic acid molecule encoding a reproductive system antigen is used in preventing, treating or ameliorating a medical condition.  
PT  
XX  
XX Disclosure; SEQ ID NO 6032; 1297pp + Sequence Listing; English.  
PS  
XX  
CC The present invention provides the protein and coding sequences of a number of human reproductive system related antigens. These can be used in the prevention and treatment of reproductive system disorders, CC including cancer. The present sequence is a genomic sequence encoding a CC protein of the invention  
XX  
SQ Sequence 10331 BP; 2718 A; 2405 C; 2599 G; 2609 T; 0 U; 0 Other;  
Query Match 40.2%; Score 608.6; DB 4; Length 10331;  
Best Local Similarity 98.6%; Pred. No. 4e-166;  
Matches 614; Conservative 0; Mismatches 9; Indels 0; Gaps 0;  
QY 892 CAGCTCTTCAAGGGAAGGTGAAGTGTGAGCTGAGCTGAAGCTGAGGTGGCCAGGTCCGAC 951  
DB 6127 CTGTGCTTCTTCCTCGAGGTGAAGTGTGAGCTGAGCTGAGGTGGCCAGGTCCGAC 6186  
QY 952 CTGTACATCTCGGCCCAAGCCGACACTTTATTGGCAACTGTGTCTCTCTTCACTGCC 1011  
DB 6187 CTGTACATCTCGGCCCAAGCCGACACTTTATTGGCAACTGTGTCTCTCTTCACTGCC 6246  
QY 1012 TTTCTGAAGCGGAGCGGACCTCCAGGGAGGCGGTCTTCTTTTTCGGCATGGACAGG 1071  
DB 6247 TTTGTGAAGCGGAGCGGACCTCCAGGGAGGCGGTCTTCTTTTTCGGCATGGACAGG 6306  
QY 1072 CCCCCTAAGCTGCGGGACGAGTCTGATTCTGGCCGGAGCACAGACCTCTGATCCTGG 1131  
DB 6307 CCCCCTAAGCTGCGGGACGAGTCTGATTCTGGCCGGAGCACAGACCTCTGATCCTGG 6366  
QY 1132 AGGACCAAGTCTGAGCTGGTCTTCCAGCCAGGCTGGCAGCAGAGGTGCTCCGGGA 1191  
DB 6367 AGGACCAAGTCTGAGCTGGTCTTCCAGCCAGGCTGGCAGCAGAGGTGCTCCGGGA 6426  
QY 1192 TTGCAAACTCTCTTCTCAGCTGCAAGATGGAGAGAGTGCCAGGACCCCTCAAGGA 1251  
DB 6427 TTGCAAACTCTCTTCTCAGCTGCAAGATGGAGAGAGTGCCAGGACCCCTCAAGGA 6486  
QY 1252 GGGAGACGCTCCATATCCAGGGCATAGGACTTGCAGGTTCCTAGGAGCAGGATCTC 1311  
DB 6487 GGGAGACGCTCCATATCCAGGGCATAGGACTTGCAGGTTCCTAGGAGCAGGATCTC 6546  
QY 1312 CCATCGCACGCTGCTTCTGCTTCTGGGAATTTCTCACACTGGCAAAGCAGTCCAGCCT 1371  
DB 6547 CCATCGCACGCTGCTTCTGCTTCTGGGAATTTCTCACACTGGCAAAGCAGTCCAGCCT 6606  
QY 1372 CCGTCTTCTGGTCCACTCTGCTCTGAGCAGCCTGGGATGCTGAATCTTTCAGAGAGATT 1431  
DB 6607 CCGTCTTCTGGTCCACTCTGCTCTGAGCAGCCTGGGATGCTGAATCTTTCAGAGAGATT 6666  
QY 1432 TTTTATAGAGAGATTTCTATAATTTTGATCAAGGTTCATGACTATCTTAGAATCTCTGT 1491  
DB 6667 TTTTATAGAGAGATTTCTATAATTTTGATCAAGGTTCATGACTATCTTAGAATCTCTGT 6726  
QY 1492 GGTTTTGAATAATCAATTGAATTC 1514  
DB 6727 GGTTTTGAATAATCAATTGAATTC 6749  
RESULT 11  
ACH35532  
ID ACH35532 standard; cDNA; 477 BP.  
XX  
XX AC ACH35532;  
XX  
XX DT 13-OCT-2003 (first entry)  
XX  
XX DE Human endothelial cell cDNA #3665.  
XX



Query Match	29.1%;	Score 441;	DB 9;	Length 479;
Best Local Similarity	98.9%;	Pred. No. 6.9e-118;		
Matches 444;	Conservative 0;	Mismatches 5;	Indels 0;	Gaps 0;
868	GAGAGTTATGTGCGCTGAGCTCCAAACAGCTCTTTCAAAGGGAAGGTGAAGTGGTGAGCGCTG	927		
31	GAGAGTTATGTGCGCTGAGCTCCACACAGCTCTTCAAAGGGAAGGTGAAGTGGTGAGCGCTG	90		
928	AAGCCTGAGGTGGGCCAGGTCGACCTGTATACATCTCTCGGCCAAGCGCACCACTTTATTGGC	987		
91	AAGCCTGAGGTGGGCCAGGTCGACCTGTATACATCTCTCGGCCAAGCGCACCACTTTATTGGC	150		
988	AACGTGTCTCCTCTCTTCACTGCCTTTGTGAAGCGGAGCGGAGACCTCCAGGGGAGGCGG	1047		
151	AACGTGTCTCCTCTCTTCACTGCCTTTGTGAAGCGGAGCGGAGACCTCCAGGGGAGGCGG	210		
1048	TCCTCTTTCTTCGCGCATCGACAGGCGCCCTTAAGCTGCGGGACGAGTTCTGATTTCTGGCGG	1107		
211	TCCTCTTTCTTCGCGCATCGACAGGCGCCCTTAAGCTGCGGGACGAGTTCTGATTTCTGGCGG	270		
1108	GAGACACAGACCTCTGTATCTGTGAGGGACACAGAGTCTGAGCTGGTCTCTCCAGCCAGGC	1167		
271	GAGACACAGACCTCTGTATCTGTGAGGGACACAGAGTCTGAGCTGGTCTCTCCAGCCAGGC	330		
1168	CTGCGAGCCAGAGTGTCTCGGGATTCGAAACTCCTCTTCTACCTGCCAAGATGAGA	1227		
331	CTGCGAGCCAGAGTGTCTCGGGATTCGAAACTCCTCTTCTACCTGCCAAGATGAGA	390		
1228	AGAGTGCCAGGGACCCCTCAAAGGAGGAGAGCGCTCCATATCCAGGGCATAGGACTTGCA	1287		
391	AGAGTGCCAGGGACCCCTCAAAGGAGGAGAGCGCTCCATATCCAGGGCATAGGACTTGCA	450		
1288	GTTTCTTAGGAGCAGGAGCATCTCCCATC	1316		
451	GTTTCTTAAGAGCAGGAGCATCTCCCATC	479		

RESULT 13	
ADQ56824	
ID	ADQ56824 standard; DNA; 610 BP.
XX	
AC	ADQ56824;
XX	
DT	21-OCT-2004 (first entry)
XX	
DE	Novel canine microarray-related DNA sequence SeqID8126.
XX	
KW	canine microarray; drug screening; toxicity assay;
KW	environmental pollutant; cellular response; gene expression profile;
KW	toxic response; liver necrosis; fatty liver disease;
KW	protein adduct formation; hepatitis; dog; ds.
XX	
OS	Canis familiaris.
XX	
PN	WO2004063324-A2.
XX	
PD	29-JUL-2004.
XX	
PF	05-MAY-2003; 2003WO-US013853.
XX	
PR	03-MAY-2002; 2002US-0377240P.
XX	
PA	(GENE-) GENE LOGIC INC.
PA	(PFIZ ) PFIZER PROD INC.
XX	
PI	Diggins JC, Porter M, Wei T;
XX	
WPI	2004-561890/54.
XX	
DR	New isolated nucleic acid molecule, useful for drug screening and
PT	toxicity assays or for assessing the impact, including toxicity, of a
PT	compound, pharmaceutical agent or environmental pollutant on a cell or
PT	cell or

living organism.

Claim 1; SEQ ID NO 8126; 41pp; English.

This invention is related to a novel isolated canine nucleic acid sequences and the construction of canine microarrays containing a significant portion of the canine genome. The isolated canine nucleic acid sequences of the invention may be useful for drug screening and toxicity assays. The invention is therefore useful for assessing the impact, including toxicity, of a compound, pharmaceutical agent or environmental pollutant on a cell or living organism. The methods are useful for detecting genes that are up- or down-regulated in canines in a disease state. The sequences are useful as diagnostic agents or markers to detect a cellular response in a sample individually or as part of a gene expression profile. It is also useful as a target for agents that modulate gene expression or activity. The database is useful for producing electronic Northernblots that allow the user to determine the cell type or tissue in which a given gene is expressed and to allow determination of the abundance or expression level of a given gene in a particular tissue or cell. The methods are useful for determining the similarity of a toxic response to one or more individual compounds. The methods are useful for predicting at least one toxic response or the likelihood that a compound or test agent will induce various specific pathologies such as those of the liver (liver necrosis, fatty liver disease, protein adduct formation or hepatitis), those of the kidney, heart, brain or testes, or other pathologies associated with at least one of the toxins. The methods are also useful for predicting or elucidating the potential cellular pathways influenced, induced or modulated by the compound or test agent due to the similarity of the expression profile compared to the profile induced by a known toxin. The present sequence is that of a canine DNA sequence which was claimed for use during the production of a canine microarray of the invention.

Query Match	Best Local Similarity	Score	406.2	DB 13	Length	610
Matches	446	Conservative	0	Mismatches	54	Indels
171	CACCAACCTC	CATGTGTCCTTACCAAGAAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCA	230			
1	CACCAATCTC	CATGTGTCCTTACCAAGAAGTACTTCAAGCTNNAGCCCTCNAGGNNTACCA	60			
231	TCGGGTTCAT	CAGCTTGGAGGATTTATCGGAGAACTGGCACCCACCCACTGGCCCCCTGA	290			
61	TCGGGTTCAT	CAGCTTGGAGGATTTATCGGAGAACTGGCACCCACCCACTGGCCCCCTGA	120			
291	GAAGCGGTGG	CATACCTGCTTGAAGTGGAGCCACGGAAGCCACAGATAAGAGACATG	350			
121	GAGCGAGTGG	CATACCTGCTTGAAGTGGAGCCACGGAAGCCACAGATAAGAGACATG	180			
351	CCCCATGAAG	GAAGAAACCCCTTTGGCCCATCTCGGATCAGTTTCATGTGAGTTTCAA	410			
181	CCCCATGAAG	GAAGAAATCCCTTTGGCCCATCTTTGGGATCAGTTTCATGTGAGTTTCAA	240			
411	CAAGTCGGAG	CTTTTTCACAGGCATTTCCCTTCAGTGTCTTACAGAAACAAATGGAGCCA	470			
241	CAAGTCCAA	GCTTTTTCAGGCATTTCCCTTCAGTGTCTTACAGAAACAAATGGATTC	300			
471	GAGATTTTCT	CCAAAGGAAATCCCGTGTCTTCCCTGCGAGGCCCGCCAGCCAGTTCCC	530			
301	GAGATTTTCT	CCAAAGGAAATCCCGTGTCTTCCCTGCGAGGCCCGCCAGTTCCC	359			
531	CGTCTTAGA	AGAACACAGAGCCACTACAGAGATACATGTTGTCAGACGAAATGGTGAA	590			
360	CGTCTTAGA	AGAACACAGAGCCACTACAGAGATACATGTTGTCAGATGAGTGGTGAG	419			
591	GACGGAGAG	GCGCCAGATTCATGCCACCTTGTCCGGCCCTTATGTGGCATTCATCTGCG	650			
420	GACGGAGAG	GCGCCAGATTCATGCCACCTTGTCCGGCCCTTATGTGGCATTCATCTGCG	479			
651	CATTGGCTCT	GACTGGAGAA	671			



XX	Drosophila melanogaster.
OS	WO200171042-A2.
XX	
XX	27-SEP-2001.
XX	
XX	23-MAR-2001; 2001WO-US009231.
XX	
XX	23-MAR-2000; 2000US-0191637P.
PR	
PR	11-JUL-2000; 2000US-00614150.
XX	
XX	(PEKE ) PE CORP NY.
PA	
XX	
XX	Venter JC, Adams M, Li PWD, Myers EW;
XX	
XX	WPI: 2001-656860/75.
DR	
DR	P-PSDB; ABB69843.
XX	
XX	New isolated nucleic acid detection reagent for detecting 1000 or more
PT	genes from Drosophila and for elucidating cell signaling and cell-cell
PT	interactions.
XX	
XX	Claim 1; SEQ ID NO 36320; 21pp + Sequence Listing; English.
PS	
XX	
CC	The invention relates to an isolated nucleic acid detection reagent
CC	capable of detecting 1000 or more genes from Drosophila. The invention is
CC	useful in developmental biology and in elucidating cell signalling and
CC	cell-cell interactions in higher eukaryotes for the development of
CC	insecticides, therapeutics, and pharmaceutical drugs. The invention
CC	discloses genomic DNA sequences (ABU16176-ABU30511), expressed DNA
CC	sequences (ABU01840-ABU16175) and the encoded proteins (ABB57737-
CC	ABB72072). The sequence data for this patent did not form part of the
CC	printed specification, but was obtained in electronic format directly
CC	from WIPO at ftp.wipo.int/pub/published_pct_sequences
CC	
XX	
SQ	Sequence 3264 BP; 904 A; 742 C; 758 G; 860 T; 0 U; 0 Other;
	Query Match 9.9%; Score 150.4; DB 4; Length 3264;
	Best Local Similarity 50.9%; Pred. No. 1.4e-32;
	Matches 541; Conservative 0; Mismatches 476; Indels 45; Gaps 6;
QY	49 TGCATGGGGCGCTTTGGGAACCGAGCCGATCAGTCTTTGGGCTCTCTGGCATTTGCAAG 108
DB	2098 TTCAAGAGCAGCTTTGGCAACCGAGCCGACACTTCTCGGATCATTTGGCCTTCGCCAAG 2039
QY	109 CTGCTAAACGTAACCTTTGGCTGTCCCTCTTGATTTGATACAGCATCAAGCCTCTCT 168
DB	2038 GCAGCTTAAATCGCACCTCTGATCCTGCGCCGTGGTGGAGT-----ATCGTAGGGGTGAA 1985
QY	169 TTCACCAACCTGCATGTGCTCTACCAAGAGTACTTCAAGCTGGAGCCCTCCAGGCTTAC 228
DB	1984 CTGGATCCCGTCAGGTACCGTTCAACACATATTTGAAGTGGAGCCCTTGAAGGAATAC 1925
QY	229 CATCGGTCATCAGCTTTGGAGGATTTTCATGGAGAAGCTGGCACCCACCTCGCCCCCT 288
DB	1924 CATCGGTCATCACCATGGCAGATTTTCATGTGGCACTTGGCCGACGACATTTGGCCAGAA 1865
QY	289 GAGAAGCGGTGGCATACTGCTTTGAGG-----TGGCAGCCCAAGCGAAGC 333
DB	1864 TCGGAGCGAGTGTCTATTTGCTACAAGGAACGATATAGCCCTTCAGCAGGAGAGAAGACGAT 1805
QY	334 CCAGATAAGAAGACGTGCCCCCATGAAGGAAGGAACCCCTTTGGCCCATTTCTGGATCAG 393
DB	1804 CCAGACNAGCCCAATTTGCCACGCCAAGGATGGCAATCTTTTGGTCCCTTTTGGGACACT 1745
QY	394 TTTTCATGTGAGTTTCAACAAGTCGGAGCTTTTACAGGCAATTTCTCTTCAGTGTCTTCTAC 453
DB	1744 TTTTCACATGACTTTTGGTCGCTCAGAGTCTCTATGCGGCACCTTCTATTTGATGTGCATCAT 1685
QY	454 AG-----AGAACATGGAGCCGAGATTTTCTCCAAAGGAACATCCGGTGTCTTGC 504
DB	1684 AGCAACGAGGCTGCAAGTGGCAGACCAAAATATCTCTGCAGAACTATATCCGTAATCTCGG 1625

THIS PAGE BLANK (USPTO)





```
Qy 61 TTTGGGAACCAAGCCGATCACTTTTGGGCTCTTGGCAATTTGCAAGCTGCTAAACCGT 120
Db 61 TTTGGGAACCAAGCCGATCACTTTTGGGCTCTTGGCAATTTGCAAGCTGCTAAACCGT 120
Qy 121 ACCTTTGGCTGTCCCTCTCTTGGATGAGTACAGAGATCACAAGCCTCCTTTTACCAACCTC 180
Db 121 ACCTTTGGCTGTCCCTCTCTTGGATGAGTACAGAGATCACAAGCCTCCTTTTACCAACCTC 180
Qy 181 CATGTGCTTACACAGAGTACTTTCAGCTGGAGCCCTCCAGGCTTACCATCGGCTATC 240
Db 181 CATGTGCTTACACAGAGTACTTTCAGCTGGAGCCCTCCAGGCTTACCATCGGCTATC 240
Qy 241 AGCTTGGAGGATTTTCATGGAGAGCTGGCAACCACTGGCCCTCTGAGAGCGGGTG 300
Db 241 AGCTTGGAGGATTTTCATGGAGAGCTGGCAACCACTGGCCCTCTGAGAGCGGGTG 300
Qy 301 GCATATCTCTTGTGAGTGGAGCCGACGGAAGCCAGATGAAGAACGCTGCCCCATGAAG 360
Db 301 GCATATCTCTTGTGAGTGGAGCCGACGGAAGCCAGATGAAGAACGCTGCCCCATGAAG 360
Qy 361 GAAGGAAACCCCTTTGGCCCATTTCTGGGATCAGTTTTTCAATGAGTTTCAACAAGTCGGAG 420
Db 361 GAAGGAAACCCCTTTGGCCCATTTCTGGGATCAGTTTTTCAATGAGTTTCAACAAGTCGGAG 420
Qy 421 CTTTTTACAGGCATTTCTTTCAGTGTCTTCTACAGAGAACAAATGGAGCCAGAGATTTTCT 480
Db 421 CTTTTTACAGGCATTTCTTTCAGTGTCTTCTACAGAGAACAAATGGAGCCAGAGATTTTCT 480
Qy 481 CCAAGGAACATCCGGTGTCTTCTGCTGCGAGGAGCCAGCCAGTTCCCGCTCTAGAA 540
Db 481 CCAAGGAACATCCGGTGTCTTCTGCTGCGAGGAGCCAGCCAGTTCCCGCTCTAGAA 540
Qy 541 GAACACAGGCGCACTACAGAGTACATGTATGTGTGTCAGAGCAATGTGAGACGGGAG 600
Db 541 GAACACAGGCGCACTACAGAGTACATGTATGTGTGTCAGAGCAATGTGAGACGGGAG 600
Qy 601 GCCAGATTTATGCCCATCTTGTGCGGCCCTATGTGGCATTCATCTGGCATTTGGCTCT 660
Db 601 GCCAGATTTATGCCCATCTTGTGCGGCCCTATGTGGCATTCATCTGGCATTTGGCTCT 660
Qy 661 GACTGGAAGAACGCTGTGCTATGTCAGTGAAGACGGGATCTCAGGCTCGCACTTCATGCCC 720
Db 661 GACTGGAAGAACGCTGTGCTATGTCAGTGAAGACGGGATCTCAGGCTCGCACTTCATGCCC 720
Qy 721 TCTCCGAGTGTGTGGCTACAGCCGACAGCAGCGGCCCTCACGATGACTATGTGC 780
Db 721 TCTCCGAGTGTGTGGCTACAGCCGACAGCAGCGGCCCTCACGATGACTATGTGC 780
Qy 781 CTGCTGACCTGAAGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTGCGTGGATGCC 840
Db 781 CTGCTGACCTGAAGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTGCGTGGATGCC 840
Qy 841 CAGTGGGTACGTTGCTACTGATTCGAGAGTTATGTGCTGAGCTCCCAAGCTCTTC 900
Db 841 CAGTGGGTACGTTGCTACTGATTCGAGAGTTATGTGCTGAGCTCCCAAGCTCTTC 900
Qy 901 AAAGGGAAGGTGAGTGTGAGCTTGAAGCTGAGGTGGCCAGGTGACCTGTATCATC 960
Db 901 AAAGGGAAGGTGAGTGTGAGCTTGAAGCTGAGGTGGCCAGGTGACCTGTATCATC 960
Qy 961 CTCGGCCAAAGCCGACCACTTTATTTGGCAACTGTGTCTCTCTTCACTGCCCTTTGGAAG 1020
Db 961 CTCGGCCAAAGCCGACCACTTTATTTGGCAACTGTGTCTCTCTTCACTGCCCTTTGGAAG 1020
Qy 1021 CCGGAGCGGGAACCTCCAGGGGAGGCGCTCTTCTTCTTGGCATGACAGGCCCTCTAAG 1080
Db 1021 CCGGAGCGGGAACCTCCAGGGGAGGCGCTCTTCTTCTTGGCATGACAGGCCCTCTAAG 1080
Qy 1081 CTGCGGGAACGAGTTCTGATTTCTGGCCGAGACACAGACCTCTGTATCTTGGAGGACAG 1140
Db 1081 CTGCGGGAACGAGTTCTGATTTCTGGCCGAGACACAGACCTCTGTATCTTGGAGGACAG 1140
Qy 1141 AGTCTGAGCTGCTCTTCCAGCCAGGCGCTGGGAGCCAGAGGTGCTCCGGGATTTGCAACT 1200
```

```
RESULT 2
US-09-333-729A-2
; Sequence 2, Application US/09333729A
; Patent No. 6270987
; GENERAL INFORMATION:
; APPLICANT: Wang, Yang
; APPLICANT: Spellman, Michael W.
; TITLE OF INVENTION: O-Fucosyltransferase
; FILE REFERENCE: P1041PDI-Substitute
; CURRENT APPLICATION NUMBER: US/09/333,729A
; CURRENT FILING DATE: 1999-06-15
; PRIOR APPLICATION NUMBER: US 08/798,741
; PRIOR FILING DATE: 1997-11-26
; NUMBER OF SEQ ID NOS: 21
; SEQ ID NO 2
; LENGTH: 1514
; TYPE: DNA
; ORGANISM: Homo Sapien
US-09-333-729A-2

Query Match 100.0%; Score 1514; DB 3; Length 1514;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1514; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGCCCGCGGGCTCCTGGGACCGGCGGTTACCTGCTCTACTGCCCCCTGCATGGGCGC 60
Db 1 ATGCCCGCGGGCTCCTGGGACCGGCGGTTACCTGCTCTACTGCCCCCTGCATGGGCGC 60
Qy 61 TTTGGGAACCAAGCCGATCACTTTCTTGGGCTCTCTGGGATTTGCAAGCTGTAAACCGT 120
Db 61 TTTGGGAACCAAGCCGATCACTTTCTTGGGCTCTCTGGGATTTGCAAGCTGTAAACCGT 120
Qy 121 ACCTTGGCTGTCCCTCTCTTGGATGAGTACAGAGATCACAAGCTCCTTTTACCAACCTC 180
Db 121 ACCTTGGCTGTCCCTCTCTTGGATGAGTACAGAGATCACAAGCTCCTTTTACCAACCTC 180
Qy 181 CATGTGCTTACACAGAGTACTTTCAGCTGGAGCCCTCCAGGCTTACCATCGGCTATC 240
Db 181 CATGTGCTTACACAGAGTACTTTCAGCTGGAGCCCTCCAGGCTTACCATCGGCTATC 240
Qy 241 AGCTTGGAGGATTTTCATGGAGAGCTGGCAACCACTGGCCCTCTGAGAGCGGGTG 300
Db 241 AGCTTGGAGGATTTTCATGGAGAGCTGGCAACCACTGGCCCTCTGAGAGCGGGTG 300
Qy 301 GCATATCTCTTGTGAGTGGAGCCGACGGAAGCCAGATGAAGAACGCTGCCCCATGAAG 360
```



Qy 241 AGCTTGGAGATTTCATGGGAAGCTGGCACCCACCCACTGGCCCCCTCGAAGCGGGT 300  
Db 4476 AGCTTGGAGATTTCATGGGAAGCTGGCACCCACCCACTGGCCCCCTCGAAGCGGGT 4535  
Qy 301 GCATCTACTGCTTTTTCAGGTGGCAGCCAGCGAAGCCAGATGAAGAACGCTGGCCCCATGAAG 360  
Db 4536 GCATCTACTGCTTTTTCAGGTGGCAGCCAGCGAAGCCAGATGAAGAACGCTGGCCCCATGAAG 4595  
Qy 361 GAAGGAACCCCTTTTGGCCCACTTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420  
Db 4596 GAAGGAACCCCTTTTGGCCCACTTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 4655  
Qy 421 CTTTTCAGAGCATTTCTTCAGTCTTCTTCAGAGCAATTCAGGTCAGAGATTTTCT 480  
Db 4656 CTTTTCAGAGCATTTCTTCAGTCTTCTTCAGAGCAATTCAGGTCAGAGATTTTCT 4715  
Qy 481 CCAAGGAACATCCGCTGCTGCCCTCCAGAGCCCCAGCCAGTTCCCGTCTGTAGAA 540  
Db 4716 CCAAGGAACATCCGCTGCTGCCCTCCAGAGCCCCAGCCAGTTCCCGTCTGTAGAG 4775  
Qy 541 GAACACAGGCCACTACAGAAGTACATGTATGTCAGACGAATGGTGAAGACGGGAGAG 600  
Db 4776 GAACACAGGCCACTACAGAAGTACATGTATGTCAGACGAATGGTGAAGACGGGAGAG 4835  
Qy 601 GCCAGATTCATGCCACCTTGTCCGGCCCTATGTGGCAATTCATCTGGCATTTGGCTCT 660  
Db 4836 GCCAGATTCATGCCACCTTGTCCGGCCCTATGTGGCAATTCATCTGGCATTTGGCTCT 4895  
Qy 661 GACTGGAAGAACGCTGTGCCATGCTGAAGGACGGGACTGCGGCTCGCACTTCATGCC 720  
Db 4896 GACTGGAAGAACGCTGTGCCATGCTGAAGGACGGGACTGCGGCTCGCACTTCATGCC 4955  
Qy 721 TCTCCGAGTGTGTGGCTACAGCCGACACAGCGCCCCCTCACGATGACTATGTC 780  
Db 4956 TCTCCGAGTGTGTGGCTACAGCCGACACAGCGCCCCCTCACGATGACTATGTC 5015  
Qy 781 CTGCTGACCTGAAGAGATTCAGAGGCTGTGAAGCTCTGGGTGAGTCCGCTGGATGCC 840  
Db 5016 CTGCTGACCTGAAGAGATTCAGAGGCTGTGAAGCTCTGGGTGAGTCCGCTGGATGCC 5075  
Qy 841 CAGTCGCTACTAGTGTGCTATGATCCAGAGTATATGCTGAGTCCCAAGCTCTTC 900  
Db 5076 CAGTCGCTACTAGTGTGCTATGATCCAGAGTATATGCTGAGTCCCAAGCTCTTC 5135  
Qy 901 AAAGGAAGTGAAGTGTGAGCCTGAAGCCTGAGTGGCCCAAGTCCAGCTGATC 960  
Db 5136 AAAGGAAGTGAAGTGTGAGCCTGAAGCCTGAGTGGCCCAAGTCCAGCTGATC 5195  
Qy 961 CTCGCCCAAGCCGACCACTTTATTTGGCACTGTGTCTCTCTTCACTGCTTTGTGAAG 1020  
Db 5196 CTCGCCCAAGCCGACCACTTTATTTGGCACTGTGTCTCTCTTCACTGCTTTGTGAAG 5255  
Qy 1021 CGGGAGCGGACCTCCAGGGGAGCGCTCTTCTTCTTCCGATGACAGGCCCCCTAAG 1080  
Db 5256 CGGGAGCGGACCTCCAGGGGAGCGCTCTTCTTCTTCCGATGACAGGCCCCCTAAG 5315  
Qy 1081 CTGGGAGCAGTCTGATTCGTCGGCGGACACAGACCTCTGATCCTGGAGGACCGAG 1140  
Db 5316 CTGGGAGCAGTCTGATTCGTCGGCGGACACAGACCTCTGATCCTGGAGGACCGAG 5375  
Qy 1141 AGTCTGAGCTGCTCTTCAGCCAGCCCTGGACGCGAGGTGCTCCGGATTCGAACT 1200  
Db 5376 AGTCTGAGCTGCTCTTCAGCCAGCCCTGGACGCGAGGTGCTCCGGATTCGAACT 5435  
Qy 1201 CCTCTTCTCACTCCCAAGATGAGAGAGTGCAGGGAACCCCTCAAGGAGGAGAGCGC 1260  
Db 5436 CCTCTTCTCACTCCCAAGATGAGAGAGTGCAGGGAACCCCTCAAGGAGGAGAGCGC 5495  
Qy 1261 TCCATATCCAGGCGCATAGGACTTTCAGGTTCTTAGGAGCAGGAGCATCTCCCATCGCAC 1320  
Db 5496 TCCATATCCAGGCGCATAGGACTTTCAGGTTCTTAGGAGCAGGAGCATCTCCCATCGCAC 5555  
Qy 1321 GTGCTTCTGCTCTTCTGGGAATTTCTCACATGGCAAGCAGTCCAGCTCCGCTCTTCT 1380

Db 5556 GTGCTTCTGCTCTTCTGGGAATTTCTCACACTGGCAAGCAGTCCAGCTCCGCTCTTCT 5615  
Qy 1381 GGTCCACTGCTCTCTGAGCAGCCTGGGATGCTGAACCTTTTCAGAGAGATTTTTCATAGA 1440  
Db 5616 GGTCCACTGCTCTCTGAGCAGCCTGGGATGCTGAACCTTTTCAGAGAGATTTTTCATAGA 5675  
Qy 1441 GAGATTTCCTATAATTTTGTATCAAGGTCATGACTATCTCTAGAACTCTCTGTGGTTTTGA 1500  
Db 5676 GAGATTTCCTATAATTTTGTATCAAGGTCATGACTATCTCTAGAACTCTCTGTGGTTTTGA 5735  
Qy 1501 AAATCATTTGAATTC 1514  
Db 5736 AAATCATTTGAATTC 5749

RESULT 4  
US-08-978-741-7  
; Sequence 7, Application US/08978741  
; Patent No. 6100076  
; GENERAL INFORMATION:  
; APPLICANT: Yang Wang, Michael W. Spellman  
; TITLE OF INVENTION: O-Pucosyltransferase  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genentech, Inc.  
; STREET: 1 DNA Way  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WinPatIn (Genentech)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/978,741  
; FILING DATE: 26-No. 6100076-1997  
; CLASSIFICATION: 435  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: 08/792498  
; FILING DATE: 31  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Svoboda, Craig G.  
; REGISTRATION NUMBER: 39,044  
; REFERENCE/DOCKET NUMBER: P1041P1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650/225-1489  
; TELEFAX: 650/952-9881  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 5009 base pairs  
; TYPE: Nucleic Acid  
; STRANDEDNESS: Single  
; TOPOLOGY: Linear  
US-08-978-741-7

Query Match 95.6%; Score 1447.4; DB 3; Length 5009;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1448; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 66 GAACAGGCGCATCCTCTTGGGCTCTCTGGCATTTGGCAAGCTGCTAAACCGTACCTT 125  
Db 1 GAACAGGCGCATCCTCTTGGGCTCTCTGGGCTCTCTGGCATTTGGCAAGCTGCTAAACCGTACCTT 60  
Qy 126 GGTGTCTCTCTCTTGGATTGAGTACAGCATTCAGAGCTCTTTTACCACCTCCATGT 185  
Db 61 GGTGTCTCTCTCTTGGATTGAGTACAGCATTCAGAGCTCTTTTACCACCTCCATGT 120  
Qy 186 GTCTTACAGAGTACTTCAAGCTGGAGCCCTCCAGGCTTACATCGGGTTCATAGCTT 245  
Db 121 GTCTTACAGAGTACTTCAAGCTGGAGCCCTCCAGGCTTACATCGGGTTCATAGCTT 180



Qy 546 CAGGCCACTACAGAACTACATCGTATGTCAGACGAAATGCTGAAGACGGAGAGGCCCA 605  
Db 481 CAGGCCACTACAGAACTACATCGTATGTCAGACGAAATGCTGAAGACGGAGAGGCCCA 540  
Qy 606 GATTTCATGCCACCTTTGTCGGCCCTCATGTGGGCANTCATCTGCGCATTTGGCTCTGACTG 665  
Db 541 GATTTCATGCCACCTTTGTCGGCCCTCATGTGGGCANTCATCTGCGCATTTGGCTCTGACTG 600  
Qy 666 GAAGAACGCTGTGCCATCTCAAGACGGGACTGACAGGCTGCAAGCTTCATGGCTCTCC 725  
Db 601 GAAGAACGCTGTGCCATCTCAAGACGGGACTGACAGGCTGCAAGCTTCATGGCTCTCC 660  
Qy 726 GCAGTGTGTGGCTACAGCCGACAGACAGCGGCCCTCATGATGACTATGTGCTGCC 785  
Db 661 GCAGTGTGTGGCTACAGCCGACAGACAGCGGCCCTCATGATGACTATGTGCTGCC 720  
Qy 786 TGACCTGAAGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGTGCCTGGATGCCCAATC 845  
Db 721 TGACCTGAAGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGTGCCTGGATGCCCAATC 780  
Qy 846 GGTCTAGCTGTGCTACTGATTCGAGAGTTATGTGCTGAGCTCAACAGCTCTTCAAAGG 905  
Db 781 GGTCTAGCTGTGCTACTGATTCGAGAGTTATGTGCTGAGCTCAACAGCTCTTCAAAGG 840  
Qy 906 GAAAGTGAAGTGGTGAAGCTCAAGCCTGAGGTGGCCGAGCTGACCTGTACATCTCG 965  
Db 841 GAAAGTGAAGTGGTGAAGCTCAAGCCTGAGGTGGCCGAGCTGACCTGTACATCTCG 900  
Qy 966 CCAAGCCGACACTTTATTTGGCACTGTCTCTCTCTCAGTGCCTTTGTGAAGCGGGA 1025  
Db 901 CCAAGCCGACACTTTATTTGGCACTGTCTCTCTCTCAGTGCCTTTGTGAAGCGGGA 960  
Qy 1026 GCGGAGCTCCAGGGAGGCGCTCTCTCTCTCGGATGAGCAGGCCCTTAAGCTGCG 1085  
Db 961 GCGGAGCTCCAGGGAGGCGCTCTCTCTCTCGGATGAGCAGGCCCTTAAGCTGCG 1020  
Qy 1086 GCACGAGTCTGATTTCTGGCCGAGCACACAGACCTCTGATCTGGAGGACACAGAGTCT 1145  
Db 1021 GCACGAGTCTGATTTCTGGCCGAGCACACAGACCTCTGATCTGGAGGACACAGAGTCT 1080  
Qy 1146 GAGTGTGCTCTCCAGCGAGGCTGGAGCCAGAGGTGCTCGGATTTGCAACTCTCT 1205  
Db 1081 GAGTGTGCTCTCCAGCGAGGCTGGAGCCAGAGGTGCTCGGATTTGCAACTCTCTCT 1140  
Qy 1206 TCTCACTGCGCAAGATGGAGAGAGTGCAGAGGCCCTCAAGAGGGGAGACGCTCCAT 1265  
Db 1141 TCTCACTGCGCAAGATGGAGAGAGTGCAGAGGCCCTCAAGAGGGGAGACGCTCCAT 1200  
Qy 1266 ATCCAGGGCATAGGACTTGCAGGTTCTTAGGAGCAGGACATCTCCATCGCACGTGCT 1325  
Db 1201 ATCCAGGGCATAGGACTTGCAGGTTCTTAGGAGCAGGACATCTCCATCGCACGTGCT 1260  
Qy 1326 TTCTGCTCTTCTGGGAATTTCTCACACTGGCAAGAGCTCCAGCTCTCTCTGCTCC 1385  
Db 1261 TTCTGCTCTTCTGGGAATTTCTCACACTGGCAAGAGCTCCAGCTCTCTCTGCTCC 1320  
Qy 1386 ACTGCTCTGAGGAGCTGGAGATGCTGACTCTTCAGAGAGATTTTATATAGAGAGAT 1445  
Db 1321 ACTGCTCTGAGGAGCTGGAGATGCTGACTCTTCAGAGAGATTTTATATAGAGAGAT 1380  
Qy 1446 TTCTATAATTTTATATAAAGGTCATGACTATCTCTAGAACTCTCTGTGGTTTTTGAATC 1505  
Db 1381 TTCTATAATTTTATATAAAGGTCATGACTATCTCTAGAACTCTCTGTGGTTTTTGAATC 1440  
Qy 1506 ATTGAATTC 1514  
Db 1441 ATTGAATTC 1449

RESULT 6  
US-08-978-741-4  
; Sequence 4, Application US/08978741  
; Patent No. 6100076

GENERAL INFORMATION:  
; APPLICANT: Yang Wang, Michael W. Spellman  
; TITLE OF INVENTION: O-Fucosyltransferase  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genentech, Inc.  
; STREET: 1 DNA Way  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WinPatIn (Genentech)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/978,741  
; FILING DATE: 26-No. 610076-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/792498  
; FILING DATE: 31  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Svoboda, Craig G.  
; REGISTRATION NUMBER: 39,044  
; REFERENCE/DOCKET NUMBER: P1041P1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650/225-1489  
; TELEFAX: 650/952-9881  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1300 base pairs  
; TYPE: Nucleic Acid  
; STRANDEDNESS: Single  
; TOPOLOGY: Linear  
; US-08-978-741-4  
  
Query Match 76.8%; Score 1163.4; DB 3; Length 1300;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1164; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
Qy 1 ATGCCCGGGGCTCTGGGACCGGCGGTTACTGCTCTACTGCCCCCTGATGGGGCG 60  
Db 136 ATGCCCGGGGCTCTGGGACCGGCGGTTACTGCTCTACTGCCCCCTGATGGGGCG 195  
Qy 61 TTGGGAACACGAGCCGATCATCTTTGGGCTCTCTGGCATTTCCAAAGCTGTAAACCGT 120  
Db 196 TTGGGAACACGAGCCGATCATCTTTGGGCTCTCTGGCATTTCCAAAGCTGTAAACCGT 255  
Qy 121 ACCTTGGTGTCTCTCTCTGGATTGAGTACAGCATCAAGGCTCTTTTACCACCACTC 180  
Db 256 ACCTTGGTGTCTCTCTCTGGATTGAGTACAGCATCAAGGCTCTTTTACCACCACTC 315  
Qy 181 CATGTGCTTACAGAGTACTTCAAGTGGAGCCCTCCAGGCTTACCATCGGTCTATC 240  
Db 316 CATGTGCTTACAGAGTACTTCAAGTGGAGCCCTCCAGGCTTACCATCGGTCTATC 375  
Qy 241 AGCTTGGAGATTTTCATGAGAGCTGGGACCCACCCACTGGGCCCTCTGAGAGCGGGTG 300  
Db 376 AGCTTGGAGATTTTCATGAGAGCTGGGACCCACCCACTGGGCCCTCTGAGAGCGGGTG 435  
Qy 301 GCATACTCTTTGAGGTGGCAGCCAGGAGCCAGATAGAGAGACGTGCCCCCATGAAG 360  
Db 436 GCATACTCTTTGAGGTGGCAGCCAGGAGCCAGATAGAGAGACGTGCCCCCATGAAG 495  
Qy 361 GAAGGAACCCCTTTGGCCCATTTCTGGGATCAGTTTCAITGTAGTTTCAACAAGTCGGAG 420  
Db 496 GAAGGAACCCCTTTGGCCCATTTCTGGGATCAGTTTCAITGTAGTTTCAACAAGTCGGAG 555  
Qy 421 CTTTTCAGAGGATTTCTTCTAGTGTCTTCTACAGAGCAATGGAGCCAGAGATTTTCT 480  
Db 556 CTTTTCAGAGGATTTCTTCTAGTGTCTTCTACAGAGCAATGGAGCCAGAGATTTTCT 615





Db 1156 CGGAGCGGACCTCCAGGGAGCGGCTCTCTTCTTCGCGATGACAGGCCCCCTTAAG 1215  
Qy 1081 CTGCGGAGCAGATTCTGATTTCTGCGGAGACACAGACCTCTGTATCTGAGGAGGACAG 1140  
Db 1216 CTGCGGAGCAGATTCTGATTTCTGCGGAGACACAGACCTCTGTATCTGAGGAGGACAG 1275  
Qy 1141 AGTCTGAGCTGGTCTCTTCCAGCCAG 1165  
Db 1276 AGTCTGAGCTGGTCTCTTCCAGCCAG 1300

## RESULT 8

US-08-978-741-16  
; Sequence 16, Application US/08978741  
; Patent No. 6100076  
; GENERAL INFORMATION:  
; APPLICANT: Yang Wang, Michael W. Spellman  
; TITLE OF INVENTION: O-Fucosyltransferase  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genentech, Inc.  
; STREET: 1 DNA Way  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WinPatIn (Genentech)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/978,741  
; FILING DATE: 26-No. 6100076-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/792498  
; FILING DATE: 31  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Svoboda, Craig G.  
; REGISTRATION NUMBER: 39,044  
; REFERENCE/DOCKET NUMBER: P1041P1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650/225-1489  
; TELEFAX: 650/952-9881  
; INFORMATION FOR SEQ ID NO: 16:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1100 base pairs  
; TYPE: Nucleic Acid  
; STRANDEDNESS: Single  
; TOPOLOGY: Linear  
US-08-978-741-16

Query Match 72.7%; Score 1100; DB 3; Length 1100;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 ATGCCCGGGGCTCTGGGACCGGCGGTACCTGCTCTACTGCCCCCTGCATGGGGCGC 60  
Db 1 ATGCCCGGGGCTCTGGGACCGGCGGTACCTGCTCTACTGCCCCCTGCATGGGGCGC 60  
Qy 61 TTGGGAACAGGCGCATCTTTCTGGGCTCTCTGGCATTTCGAAAGCTGCTAAACCGT 120  
Db 61 TTGGGAACAGGCGCATCTTTCTGGGCTCTCTGGCATTTCGAAAGCTGCTAAACCGT 120  
Qy 121 ACCTTGGCTGTCCCTCTCTGGATTGAGTACAGCATCACAAAGCTCTTTTACCAACCTC 180  
Db 121 ACCTTGGCTGTCCCTCTCTGGATTGAGTACAGCATCACAAAGCTCTTTTACCAACCTC 180  
Qy 181 CATGTGCTTACAGAGTACTTCAAGTGGAGCCCTCCAGGCTTACATCGGTATC 240  
Db 181 CATGTGCTTACAGAGTACTTCAAGTGGAGCCCTCCAGGCTTACATCGGTATC 240

Qy 241 AGCTTGGAGGATTTCTATGGAAGCTGGGACCCACCCACTGGGCCCCCTGAGAAGCGGGTG 300  
Db 241 AGCTTGGAGGATTTCTATGGAAGCTGGGACCCACCCACTGGGCCCCCTGAGAAGCGGGTG 300  
Qy 301 GCATCTCTCTTTGAGGTGGCAGCCAGGAGGAGCCAGATAAGAAAGACGTGCCCCCATGAAG 360  
Db 301 GCATCTCTCTTTGAGGTGGCAGCCAGGAGGAGCCAGATAAGAAAGACGTGCCCCCATGAAG 360  
Qy 361 GAAGGAACCCCTTTGGGCCATTTCTGGGATCAGTTTCTATGTGAGTTTCAACAAGTCGGAG 420  
Db 361 GAAGGAACCCCTTTGGGCCATTTCTGGGATCAGTTTCTATGTGAGTTTCAACAAGTCGGAG 420  
Qy 421 CTTTTTACAGGCATTTCTTCTAGTGTCTTCTACAGAGACAAATGGAGCCAGAGATTTCT 480  
Db 421 CTTTTTACAGGCATTTCTTCTAGTGTCTTCTACAGAGACAAATGGAGCCAGAGATTTCT 480  
Qy 481 CCAAAGGAACATCCGGTGTCTTCCCTGCGAGGAGCCCGAGCCAGTTCCCTCTCTAGAA 540  
Db 481 CCAAAGGAACATCCGGTGTCTTCCCTGCGAGGAGCCCGAGCCAGTTCCCTCTCTAGAA 540  
Qy 541 GAAACACAGGCCACTTACAGAAATGATGTATGGTTCAGACGAAATGGTGAAGACGGAGAG 600  
Db 541 GAAACACAGGCCACTTACAGAAATGATGTATGGTTCAGACGAAATGGTGAAGACGGAGAG 600  
Qy 601 GCCCAGATTCTATGCCCACTTGTTCGGGCCCTATGTGGGCATTCATCTCGGCATTGGCTCT 660  
Db 601 GCCCAGATTCTATGCCCACTTGTTCGGGCCCTATGTGGGCATTCATCTCGGCATTGGCTCT 660  
Qy 661 GACTGGAAGAACCGCTGTGCTATGCTGGAAGGACGGGATCTGAGGCTCGCACTTCATGGCC 720  
Db 661 GACTGGAAGAACCGCTGTGCTATGCTGGAAGGACGGGATCTGAGGCTCGCACTTCATGGCC 720  
Qy 721 TCTCCGCAAGTGTGGGCTACAGCCGACGACAGCGGCCCTCCACGATGACTATGTGC 780  
Db 721 TCTCCGCAAGTGTGGGCTACAGCCGACGACAGCGGCCCTCCACGATGACTATGTGC 780  
Qy 781 CTGCTGACCTGGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGCTGAGTCTCGATGCC 840  
Db 781 CTGCTGACCTGGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGCTGAGTCTCGATGCC 840  
Qy 841 CAGTCGCTTACGTTGCTACTGATTCGAGAGTTATGTGCTGAGCTCCAAAGCTCTTC 900  
Db 841 CAGTCGCTTACGTTGCTACTGATTCGAGAGTTATGTGCTGAGCTCCAAAGCTCTTC 900  
Qy 901 AAAGGAAAGGTGAAGGTGAGGCTTGAAGGCTGAGGTGGGCCCGAGCTCGACCTGTACATC 960  
Db 901 AAAGGAAAGGTGAAGGTGAGGCTTGAAGGCTGAGGTGGGCCCGAGCTCGACCTGTACATC 960  
Qy 961 CTCGGCCAAAGCCGACCACTTTATTGGCAACTGTGTCTCTCTTCACTGCTTTGTGAAG 1020  
Db 961 CTCGGCCAAAGCCGACCACTTTATTGGCAACTGTGTCTCTCTTCACTGCTTTGTGAAG 1020  
Qy 1021 CGGAGCGGAGCTTCCAGGGAGGCGCTCTTCTTCTTCGCGATGACAGGCCCTCTAAG 1080  
Db 1021 CGGAGCGGAGCTTCCAGGGAGGCGCTCTTCTTCTTCGCGATGACAGGCCCTCTAAG 1080  
Qy 1081 CTGCGGAGCAGGTTCTGATT 1100  
Db 1081 CTGCGGAGCAGGTTCTGATT 1100

RESULT 9  
US-09-333-729A-16  
; Sequence 16, Application US/09333729A  
; Patent No. 6270987  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Yang  
; APPLICANT: Spellman, Michael W.  
; TITLE OF INVENTION: O-Fucosyltransferase  
; FILE REFERENCE: P1041P1D1-Substitute  
; CURRENT APPLICATION NUMBER: US/09/333,729A  
; CURRENT FILING DATE: 1999-06-15  
; PRIOR APPLICATION NUMBER: US 08/798,741



; PRIOR FILING DATE: 1997-11-26  
; NUMBER OF SEQ ID NOS: 21  
; SEQ ID NO 16  
; LENGTH: 1100  
; TYPE: DNA  
; ORGANISM: Homo Sapien  
US-09-333-729A-16

Query Match 72.7%; Score 1100; DB 3; Length 1100;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCCCCGGGCTCTCGGACCCCGCGTTACCTCTACTGCCCCCTGCAATGGGGCG 60  
Db 1 ATGCCCCGGGCTCTCGGACCCCGCGTTACCTCTACTGCCCCCTGCAATGGGGCG 60

QY 61 TTTGGGAACCGGCGATCACTTCTTGGGCTCTGGCATTTGCAAGCTGCTAAACGTT 120  
Db 61 TTTGGGAACCGGCGATCACTTCTTGGGCTCTGGCATTTGCAAGCTGCTAAACGTT 120

QY 121 ACCTTGGGTGCTCCTCTTGGATTGAGTACAGCATCAAGCCTCTTTTACCAACCTC 180  
Db 121 ACCTTGGGTGCTCCTCTTGGATTGAGTACAGCATCAAGCCTCTTTTACCAACCTC 180

QY 181 CATGTCTCTACCAAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGTCTATC 240  
Db 181 CATGTCTCTACCAAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGTCTATC 240

QY 241 AGCTTGGAGGATTTGATGAGAGTGGGACCCACCTGCCCCCTTGAGAGCGGGTG 300  
Db 241 AGCTTGGAGGATTTGATGAGAGTGGGACCCACCTGCCCCCTTGAGAGCGGGTG 300

QY 301 GCATCTCTTTGAGTGGGACCCAGCGAGCCAGATAGAGACGTCGCCCATGAAG 360  
Db 301 GCATCTCTTTGAGTGGGACCCAGCGAGCCAGATAGAGACGTCGCCCATGAAG 360

QY 361 GAAGAAACCCCTTTGGGCCATTTCTGGGATCAGTTTTCATGTGATTTTCAAGTCGGAG 420  
Db 361 GAAGAAACCCCTTTGGGCCATTTCTGGGATCAGTTTTCATGTGATTTTCAAGTCGGAG 420

QY 421 CTTTTTACAGGATTTCTTCACTGCTTCTACAGAGAAATGATGAGCAAGATTTCT 480  
Db 421 CTTTTTACAGGATTTCTTCACTGCTTCTACAGAGAAATGATGAGCAAGATTTCT 480

QY 481 CCAAGGAACATCCGGTCTGCTGCGCCAGAGCCAGAGCCAGTCCCGTCTCTAGAA 540  
Db 481 CCAAGGAACATCCGGTCTGCTGCGCCAGAGCCAGAGCCAGTCCCGTCTCTAGAA 540

QY 541 GAACACAGGCCACTACAGAGTACATGTTGATGATGATGATGATGATGATGATGATG 600  
Db 541 GAACACAGGCCACTACAGAGTACATGTTGATGATGATGATGATGATGATGATGATG 600

QY 601 GCCAGATTCATGCCCACTTGTCCGGCCCTATGTGGGATTCATCTCGCATTTGGTCT 660  
Db 601 GCCAGATTCATGCCCACTTGTCCGGCCCTATGTGGGATTCATCTCGCATTTGGTCT 660

QY 661 GACTGGAAGACGCTGTGCACTGTGAGGACGGGCTGAGGCTCCGACTTCTATGCGC 720  
Db 661 GACTGGAAGACGCTGTGCACTGTGAGGACGGGCTGAGGCTCCGACTTCTATGCGC 720

QY 721 TCTCCGAGTGTGGGCTACAGCCGACAGCGGCCCTCCATGACTATGTGC 780  
Db 721 TCTCCGAGTGTGGGCTACAGCCGACAGCGGCCCTCCATGACTATGTGC 780

QY 781 CTGCTGACCTGAAGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGTCTGATGCC 840  
Db 781 CTGCTGACCTGAAGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGTCTGATGCC 840

QY 841 CAGTCGGTCTAGTTGCTACTGATTCAGAGGATTTATGTCCTGAGCTCCACAGCTTTC 900  
Db 841 CAGTCGGTCTAGTTGCTACTGATTCAGAGGATTTATGTCCTGAGCTCCACAGCTTTC 900

QY 901 AAAGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGTGGGCCGAGCTCGACCTGTACATC 960

Db 901 AAAGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGTGGCCAGGCTCGACCTGTACATC 960

QY 961 CTGSCCAAGCGGACCACTTTATTGGCAACTGTGTCTCTCTTCACTGCGCTTTGTGAAG 1020  
Db 961 CTGSCCAAGCGGACCACTTTATTGGCAACTGTGTCTCTCTTCACTGCGCTTTGTGAAG 1020

QY 1021 CGGAGCGGGACCTCCAGGGAGGCGCTTTCTTTCTCGGCATGACAGGCCCTTAAG 1080  
Db 1021 CGGAGCGGGACCTCCAGGGAGGCGCTTTCTTTCTCGGCATGACAGGCCCTTAAG 1080

QY 1081 CTGCGGACGAGTTCTGATT 1100  
Db 1081 CTGCGGACGAGTTCTGATT 1100

RESULT 10  
US-09-270-767-14353  
; Sequence 14353, Application US/09270767  
; Patent No. 6703491  
; GENERAL INFORMATION:  
; APPLICANT: Homburger et al.  
; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*  
; FILE REFERENCE: File Reference: 7326-094  
; CURRENT APPLICATION NUMBER: US/09/270,767  
; CURRENT FILING DATE: 1999-03-17  
; NUMBER OF SEQ ID NOS: 62517  
; SOFTWARE: Patent in ver. 2.0  
; SEQ ID NO 14353  
; LENGTH: 1320  
; TYPE: DNA  
; ORGANISM: *Drosophila melanogaster*  
US-09-270-767-14353

Query Match 11.1%; Score 168; DB 4; Length 1320;  
Best Local Similarity 51.6%; Pred. No. 6.1e-41;  
Matches 564; Conservative 0; Mismatches 485; Indels 45; Gaps 6;

QY 17 GGGACCCGGCGTTACTGCTCTACTGCCCTGCGATGGGGCGCTTTGGGAACCGGCG 76  
Db 73 GCGATCCCAATGGGTACCTCACCTACTGCTCGTGTATGGGACGCTTTGGCAACCGGCG 132

QY 77 ATCACTTCTTGGGCTCTCTGGCATTTGCAAGCTGCTAAACCGTACCTTGGCTGCTCCTC 136  
Db 133 ACCACTTCTTGGGATCATTTGGCTTGGCAAGCGCTTAATGCGACCTGATCTGCGCG 192

QY 137 CTTGGATTGATACAGCATCACAGGCTCTTTTACCAACCTCCATGCTGTCTACAGA 196  
Db 193 CGTGGTGGAGT-----ATCGTAGGGTGAACCTGCGATCCCGTCAGGTACCGTTCAACA 246

QY 197 AGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGTCATCAGCTTGGAGGATTTCA 256  
Db 247 CATATTTTGAAGTGGAGCCCTTGAAGGAATACCATCGCTCATCACCATGGCAGATTTCA 306

QY 257 TGGAGAGCTGGCACCCACCCACTGGCCCTCTGAGAGCGGTGGCTACTGCTTTGAGG 316  
Db 307 TGTGCACTGGCGACGACATTTTGGCCAGAAATCGAGCGAGTGTCTATTTGTCTACAAG 366

QY 317 -----TGGCAGCCCGAGGAAAGCCAGATGAAGAAGCGTSCCCCATGAAGG 361  
Db 367 AACGATATAGCTTTTTCAGGAGAAAGACGATCAGACAGCCCAATTTGCCACGCCAAGG 426

QY 362 AAGAAACCCCTTTGGCCCATTTGGGATCAGTTTTCATGTAGTTTCAACAGTGGAGC 421  
Db 427 ATGCAATCTCTTTTGGTCCCTTTTGGGACACTTTTTCATAGACTTTTGTGGGTGAGT 486

QY 422 TTTTTCAGGCACTTTCTTTCAGTCTTCTCTACAG-----AGAAACAATGAGGCCAGA 472  
Db 487 TCTATGCGCACTTCATTTTGTATGTGATCATAGAACAGGCTGCCAAGTGGCAGACA 546

QY 473 GATTTCTTCCAAAGGAACATCCGGTGTCTTGGCTTCCAGAGGAGCCCGACCGATTTCCCG 532  
Db 547 AATATCTGCGAGATCATATCCCGTACTCGCGTTTCCCGGAGCTCCGCTAGTTTCTCTG 606

```
Qy 533 TCCTAGAGAACACAGGCCACTACAGAAAGTACATGGTATGGTCAGACGAAATGGTGAAGA 592
Db 607 TTCAGCTAGAGAACTGCAAGCTGCGAGCGCTACTTGCAGTGGAGTCAACGGTATAGGGAG 666
Qy 593 CGGAGAGGCCAGATTCATGCCCCACCTTGTCCGCC---CCTATGTGGGCATTCATCTGC 649
Db 667 CATCTAAGGAATTTTCATCGAGAGCAGTGTGCCTCGGGGTGCTTTTGGGCGCATTCATCTGC 726
Qy 650 GCATTGGCTCTGACTGGAGAAACGCTGTGCATCTGTGAAGGACGG-----GACTG 700
Db 727 GCAACGGTATCATGTTGGGTGAGAGCCTGTGAGACGCTCAGGATAGCCAGCATCTGTTTG 786
Qy 701 CAGGCTCGCACTTCATGGCTCTCCGAGTGTGTGGCTACAGCCGACGACAGGGGCC 760
Db 787 CTGCGCCGCACTGTGGGTATATAAATGAACGTGGTGCACTTACCGGAGCTCTTGCA 846
Qy 761 CCCTCAGCATGACTATGTGCTGCTGACCTGAGAGGAGATCCAGAGGGCTGTGAAGCTCT 820
Db 847 TGCCCTCCAAGGAGGGATCATCCGCCAGCTAAAGAGAACCATTAAGACAGTGGGCCAAA 906
Qy 821 GGGTGAGTGTGCTGGATGCCAGTCCGCTCTACGTGTGCTACTGATTCGAGAGTATATGTC 880
Db 907 CTCAGCCGCAACAGAAATCAATCAGTTTTTCGTGGGCTCAGACTCCAATCAGATGTTG 966
Qy 881 CTGAGCTCCACAGCTCTTCAAGGGAAG---TGAAGTGTGTGAGCTGTGAGCTCGAG 937
Db 967 GTGAACATAACACGGCCCTTAGTCCATGGGCATCAGTGTGCAAGCTGACGGAGATG 1026
Qy 938 TGGCCAGGTGCACTGTATACCTTCGGCGAAGCCGACCACTTTATTTGGCAACTGTGTCT 997
Db 1027 ATCTTACCTTGGACTTGGCCATCTTCGACACAGTCCGACCACTTATTCGGCAACTGTAT 1086
Qy 998 CTTCTTCACTGCTTTGTGAAGCGGAGCGGACCTCCAGGGGGAGCGCTCTCTTTCT 1057
Db 1087 CTTCTTACTCGGCATCTGTAAGAGGGAACGAGATGTGCACGGTTTTCATCGTACTTCT 1146
Qy 1058 TCGGCATGACAGG 1071
Db 1147 GGGGATCCCCAAG 1160

RESULT 11
US-09-270-767-30507
; Sequence 30507, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 30507
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-30507

Query Match 4.1%; Score 62.2; DB 4; Length 722;
Best Local Similarity 49.5%; Pred. No. 1.7e-08;
Matches 225; Conservative 0; Mismatches 218; Indels 12; Gaps 2;

Qy 629 CCTATGTGGGCAATTCATCTGCGCATTTGGCTCTGACTGGAAGAACGCTGTGCCATGTCTGA 688
Db 108 CTTTTTGGGCAATTCATCTGCGCAACGGTATCGATTGGGTGAGAGCCTGTGAGCACGTCA 167
Qy 689 AGGACGG-----GACTCGAGGCTCGCACTTCATGGCTCTCCGAGTGTGGGCT 739
Db 168 AGGATAGCCAGCATCTGTTTGTCTCGCGCAGTGTCTGGGCTATATAAATGAACGTGTG 227
Qy 740 ACAGCCGACGACAGCGGCCCCCTCAGCATGACTATGTGCTGCTGACCTGAAGGAGA 799
```

```
Db 228 CACTTACCCGGAGCTCTGCATGCCCCCTCAAGAGGAGCGCATATCCGCCAGCTTAAAGAA 287
Qy 800 TCCAGAGGGCTGTGAAGTCTTGGGTGAGGTGCGTGGATGCCAGTCCGCTCTACGTTGCTA 859
Db 288 CCATTTAAGAACGTGCGCCAAACTCAGCCCGGACAAACGAAATCAAATCAGTCTTTCGTGGCGT 347
Qy 860 CTGATTTCCGAGAGTTATGTGCTGAGCTCCAAACAGCTCTTCAAGAGGAAG---TGAAGG 916
Db 348 CAGACTCCAATCACATGATTTGGTGAACATAACACGGCCCTTAGTCGATGGCATCAGTG 407
Qy 917 TGGTGAGCTTGAAGCCTGAGGTGGCCAGGTGCGACCTGTACATCTCGGCCAAGCCGACC 976
Db 408 TCCACAGCTGACGGAGGATGATCTTACCTGAGTCTGGCCATTTCCGACAGTGAACC 467
Qy 977 ACTTTATTTGGCACTGTGTCTCTCTCACTGCTTGTGAAGCGGAGCGGAGCTCC 1036
Db 468 ACTTTATCGGCACTGTATATCTCTTACTCGGCATTCGTAAGAGGAAACGAGATGTGC 527
Qy 1037 AGGGAGGCCGCTCTTCTTTCTTCGGCATGGACAGG 1071
Db 528 ACGGTTTTCCATCGTACTTCTGGGATTCGCCAAG 562

RESULT 12
US-09-270-767-28905/c
; Sequence 28905, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 28905
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-28905

Query Match 3.9%; Score 58.6; DB 4; Length 998;
Best Local Similarity 60.2%; Pred. No. 2.7e-07;
Matches 97; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

Qy 911 TGAAGTGTGTGAGCTGGAAGCCTGAGGTGGCCAGGTGCGACCTGTACATCTCGGCCAAG 970
Db 896 TCAGTGTGCACAAGCTGACGGAGGATGATCTTACCTGGACCTTGGCCATTCGACAGT 837
Qy 971 CGACCACTTTTATTTGGCAACTGTGTCTCTCTCACTGCTTGTGAAGCGGAGCGGG 1030
Db 836 CGAACCACTTTTATCGGCACTGTATATCTCTTACTCGGCATTCGTAAGAGGAACGAG 777
Qy 1031 ACCTCCAGGGAGCGCTCTCTTCTTCGCGCATGGACAGG 1071
Db 776 ATGTGCACGGTTTTTCATCTGACTTCTGGGATTCGCCAAG 736

RESULT 13
US-09-270-767-13022/c
; Sequence 13022, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13022
; LENGTH: 2034
```

```

RESULT 15
US-09-949-016-16285/c
; Sequence 16285, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012

```

TWO PAGE BLANK (USPTO)

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 9, 2005, 08:00:04 ; Search time 1355.3 Seconds  
(without alignments)  
9238.165 Million cell updates/sec

Title: US-09-774-954-1  
Perfect score: 1514  
Sequence: 1 ATGCCCGCGGCTCTGGCA.....TTTTGAAAATCAATGAATTC 1514

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 9794790 seqs, 4134909567 residues

Total number of hits satisfying chosen parameters: 19589580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :				Published Applications NA:*	
1:	/cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*				
2:	/cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*				
3:	/cgn2_6/ptodata/1/pubpna/US05_NEW_PUB.seq.*				
4:	/cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*				
5:	/cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*				
6:	/cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*				
7:	/cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*				
8:	/cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*				
9:	/cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*				
10:	/cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*				
11:	/cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*				
12:	/cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*				
13:	/cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*				
14:	/cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*				
15:	/cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*				
16:	/cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*				
17:	/cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq.*				
18:	/cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq.*				
19:	/cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq.*				
20:	/cgn2_6/ptodata/1/pubpna/US10G_PUBCOMB.seq.*				
21:	/cgn2_6/ptodata/1/pubpna/US10H_PUBCOMB.seq.*				
22:	/cgn2_6/ptodata/1/pubpna/US10I_PUBCOMB.seq.*				
23:	/cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*				
24:	/cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*				
25:	/cgn2_6/ptodata/1/pubpna/US11A_PUBCOMB.seq.*				
26:	/cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*				
27:	/cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*				
28:	/cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*				

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1514	100.0	1514	11	US-09-774-954-1
2	1512.4	99.9	11284	11	US-09-774-954-5
3	1510.8	99.8	5218	16	US-10-301-822-168
4	1510.8	99.8	5218	22	US-10-956-157-4752
5	1447.4	95.6	5009	11	US-09-774-954-7
					Sequence 1, Appli
					Sequence 5, Appli
					Sequence 168, App
					Sequence 4752, Ap
					Sequence 7, Appli

6	1163.4	76.8	1300	11	US-09-774-954-4	Sequence 4, Appli
7	1100	72.7	1100	11	US-09-774-954-16	Sequence 16, Appl
8	608.6	40.2	10331	10	US-09-764-891-6032	Sequence 6032, Ap
9	471.4	31.1	1400	22	US-10-956-157-9986	Sequence 9986, Ap
10	471.4	31.1	1506	22	US-10-956-157-4751	Sequence 4751, Ap
11	442.8	29.2	477	10	US-09-918-995-22744	Sequence 22744, A
12	441	29.1	479	10	US-09-918-995-19437	Sequence 19437, A
13	166.4	11.0	1209	26	US-11-097-143-18161	Sequence 18161, A
14	150.4	9.9	3264	26	US-11-097-143-18160	Sequence 18160, A
15	114.6	7.6	402	9	US-09-783-590-11501	Sequence 11501, A
16	108.6	7.2	3793	26	US-11-097-143-18082	Sequence 18082, A
17	71.8	4.7	3567	26	US-11-097-143-18169	Sequence 18169, A
18	60	4.0	60	10	US-09-908-975-17287	Sequence 17287, A
19	56	3.7	760	24	US-10-820-474A-236	Sequence 236, App
20	55.6	3.7	521	13	US-09-925-065A-60244	Sequence 60244, A
21	42.4	2.8	1438	18	US-10-282-122A-19431	Sequence 19431, A
22	41.2	2.7	114793	16	US-10-148-806-3	Sequence 3, Appli
23	41.2	2.7	114793	24	US-10-859-792-3	Sequence 3, Appli
24	40.6	2.7	2036	18	US-10-369-493-43154	Sequence 43154, A
25	40.4	2.7	1212	21	US-10-425-115-114097	Sequence 114097, A
26	40.4	2.7	5932	16	US-10-205-823-312	Sequence 312, App
27	40.4	2.7	5932	26	US-11-051-454-312	Sequence 312, App
28	40.2	2.7	726	20	US-10-437-963-19394	Sequence 19394, A
29	40	2.6	40	11	US-09-774-954-10	Sequence 10, Appl
30	40	2.6	40	11	US-09-774-954-12	Sequence 12, Appl
31	40	2.6	2032	20	US-10-437-963-8098	Sequence 8098, Ap
32	39.2	2.6	618	13	US-09-925-065A-571171	Sequence 571171, A
33	39	2.6	372	18	US-10-282-122A-31476	Sequence 31476, A
34	39	2.6	1209	10	US-09-948-820-37	Sequence 37, Appl
35	39	2.6	1209	21	US-10-613-076-37	Sequence 37, Appl
36	39	2.6	1555	24	US-10-220-335-120	Sequence 120, App
37	39	2.6	1575	19	US-10-072-012-131	Sequence 131, App
38	39	2.6	1654	24	US-10-220-335-121	Sequence 121, App
39	39	2.6	1963	24	US-10-220-335-464	Sequence 464, App
40	39	2.6	1963	24	US-10-220-335-465	Sequence 465, App
41	38.6	2.5	407	9	US-09-960-352-11732	Sequence 11732, A
42	38.6	2.5	2358	10	US-09-919-039-190	Sequence 190, App
43	38.6	2.5	2384	22	US-10-278-698-123	Sequence 123, App
44	38.6	2.5	2384	22	US-10-278-698-638	Sequence 638, App
45	38.6	2.5	2670	17	US-10-247-671-64	Sequence 64, Appl

ALIGNMENTS

RESULT 1  
US-09-774-954-1  
; Sequence 1, Application US/09774954  
; Publication No. US20040241645A1  
; GENERAL INFORMATION:  
; APPLICANT: Yang Wang, Michael W. Spellman  
; TITLE OF INVENTION: O-Fucosyltransferase  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genentech, Inc.  
; STREET: 1 DNA Way  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WinPatIn (Genentech)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/774,954  
; FILING DATE: 30-Jan-2001  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/978,741  
; FILING DATE: 26-NOV-1997  
; APPLICATION NUMBER: 08/792,498

```
; FILING DATE: 31-JAN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Svoboda, Craig G.
; REGISTRATION NUMBER: 39,044
; REFERENCE/DOCKET NUMBER: P1041P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-1489
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1514 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-774-954-1

Query Match      100.0%; Score 1514; DB 11; Length 1514;
Best Local Similarity 100.0%; Pred No. 0;
Matches 1514; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCCCGGGCTCTGGGACCGGCGGTTACCTGCTCTACTGCCCCCTGCATGGGGCC 60
DB 1 ATGCCCGGGCTCTGGGACCGGCGGTTACCTGCTCTACTGCCCCCTGCATGGGGCC 60
QY 61 TTGGGAAACAGGCGCATCACTTCTGGGCTCTCTGCGATTTGCAAGCTGCTAAACCGT 120
DB 61 TTGGGAAACAGGCGCATCACTTCTGGGCTCTCTGCGATTTGCAAGCTGCTAAACCGT 120
QY 121 ACTTGGCTGTCCTCTCTGGATTTAGTACCAAGCATCAAGGCTCTCTTTCACCAACCTC 180
DB 121 ACTTGGCTGTCCTCTCTGGATTTAGTACCAAGCATCAAGGCTCTCTTTCACCAACCTC 180
QY 181 CATGTGCTTACAGAACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGTCATC 240
DB 181 CATGTGCTTACAGAACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGTCATC 240
QY 241 AGCTTGGAGATTTTCATGAGAGCTGGCCACCCACCTGCGCCCTCAGAAAGCGGTG 300
DB 241 AGCTTGGAGATTTTCATGAGAGCTGGCCACCCACCTGCGCCCTCAGAAAGCGGTG 300
QY 301 GCATCTGCTTTGAGTGGCGCCAGCGAGCCGACATGAAGACGTCGCCCATGAG 360
DB 301 GCATCTGCTTTGAGTGGCGCCAGCGAGCCGACATGAAGACGTCGCCCATGAG 360
QY 361 GAAGGAAACCCCTTTGGCCCATTTCTGGGATCAGTTTCAATGAGTTTCAACAGTCGGAG 420
DB 361 GAAGGAAACCCCTTTGGCCCATTTCTGGGATCAGTTTCAATGAGTTTCAACAGTCGGAG 420
QY 421 CTTTTCACGCAATTCCTTTCAGTGGTTCTTACAGAGAACAATGGAGCCAGAGATTTTCT 480
DB 421 CTTTTCACGCAATTCCTTTCAGTGGTTCTTACAGAGAACAATGGAGCCAGAGATTTTCT 480
QY 481 CCAAGGAAACATCCGGTGTCTTGGCCCTGCGAGGAGCCCGCCAGTTCCCGTCTCTAGAA 540
DB 481 CCAAGGAAACATCCGGTGTCTTGGCCCTGCGAGGAGCCCGCCAGTTCCCGTCTCTAGAA 540
QY 541 GAACACAGGCACTACAGAACTACATGTTGTTGTCAGAGAAATGGTGAAGACGGGAG 600
DB 541 GAACACAGGCACTACAGAACTACATGTTGTTGTCAGAGAAATGGTGAAGACGGGAG 600
QY 601 GCCAGATTCATGCCACCTTCTGGGCGCCCTATGTTGGGCAATTCATCTGGCAATTTGGCT 660
DB 601 GCCAGATTCATGCCACCTTCTGGGCGCCCTATGTTGGGCAATTCATCTGGCAATTTGGCT 660
QY 661 GACTGGAAGACGCTTGGCCATCTGAGGACGGGACTGCGAGGCTCGCACTTCATGGCC 720
DB 661 GACTGGAAGACGCTTGGCCATCTGAGGACGGGACTGCGAGGCTCGCACTTCATGGCC 720
QY 721 TCTCCGCAAGTGTGGGCTACAGCGCGACGACAGCGGCGCCCTCTACGATGATATGTC 780
DB 721 TCTCCGCAAGTGTGGGCTACAGCGCGACGACAGCGGCGCCCTCTACGATGATATGTC 780
```

```
QY 781 CTGCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGTCCGCTGGATGCC 840
DB 781 CTGCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGTCCGCTGGATGCC 840
QY 841 CAGTCGGTCTACGTTGCTACTGATTCGAGAGTTATATGCTCGAGCTCCAAACAGCTCTTC 900
DB 841 CAGTCGGTCTACGTTGCTACTGATTCGAGAGTTATATGCTCGAGCTCCAAACAGCTCTTC 900
QY 901 AAAGGGAAGGTGAAGGTGTGAGCCTTGAAGCCTTGAGGTGGCCAGGCTCGACCTGTATCATC 960
DB 901 AAAGGGAAGGTGAAGGTGTGAGCCTTGAAGCCTTGAGGTGGCCAGGCTCGACCTGTATCATC 960
QY 961 CTCGGCCAAAGCCGACCACTTTATTGGCAACTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1020
DB 961 CTCGGCCAAAGCCGACCACTTTATTGGCAACTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1020
QY 1021 CGGAGAGGGGACCTCCAGGGGAGGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1080
DB 1021 CGGAGAGGGGACCTCCAGGGGAGGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1080
QY 1081 CTGCGGAGCGAGTTCTGATTTCTGGCGGAGCAGCAGGACCTCTGATCTCTGGAGGAGCCAG 1140
DB 1081 CTGCGGAGCGAGTTCTGATTTCTGGCGGAGCAGCAGGACCTCTGATCTCTGGAGGAGCCAG 1140
QY 1141 AGTCTGAGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1200
DB 1141 AGTCTGAGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1200
QY 1201 CCTCTTCTCACCTGCGCAAGATGGAGAGGTGCGAGGACCCCTCAAGGAGGAGAGCGC 1260
DB 1201 CCTCTTCTCACCTGCGCAAGATGGAGAGGTGCGAGGACCCCTCAAGGAGGAGAGCGC 1260
QY 1261 TCCATATCCAGGAGCATAGGACTTTCAGGCTTCTAGGAGCAGGAGCATCTCCCATCGCAC 1320
DB 1261 TCCATATCCAGGAGCATAGGACTTTCAGGCTTCTAGGAGCAGGAGCATCTCCCATCGCAC 1320
QY 1321 GTGCTTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1380
DB 1321 GTGCTTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1380
QY 1381 GGTCCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1440
DB 1381 GGTCCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1440
QY 1441 GAGATTTCTATATTTGATACAGGTCATGACTATCTAGAACTCTCTGTTGTTTGA 1500
DB 1441 GAGATTTCTATATTTGATACAGGTCATGACTATCTAGAACTCTCTGTTGTTTGA 1500
QY 1501 AAATCATTGAATTC 1514
DB 1501 AAATCATTGAATTC 1514
```

```
RESULT 2
US-09-774-954-5
; Sequence 5, Application US/09774954
; Publication No. US20040241645A1
; GENERAL INFORMATION:
; APPLICANT: Yang Wang, Michael W. Spellman
; TITLE OF INVENTION: O-Fucosyltransferase
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESS: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
```

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/774,954  
FILING DATE: 30-Jan-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/978,741  
FILING DATE: 26-Nov-1997  
APPLICATION NUMBER: 08/792,498  
FILING DATE: 31-Jan-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Svoboda, Craig G.  
REGISTRATION NUMBER: 39,044  
REFERENCE/DOCKET NUMBER: P1041P1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650/225-1489  
TELEFAX: 650/952-9881  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 11284 base pairs  
TYPE: Nucleic Acid  
STRANDEDNESS: Single  
TOPOLOGY: Linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 5:  
US-09-774-954-5

Query Match 99.9%; Score 1512.4; DB 11; Length 11284;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1513; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	1	ATGCCCGGGGCTCTGGACCCGCGGTACCTGCTCTACTGCGCCCTGCATGGGGGC	60
DB	4236	ATGCCCGGGGCTCTGGACCCGCGGTACCTGCTCTACTGCGCCCTGCATGGGGGC	4295
QY	61	TTTGGGAACACGCGCGATCACTTCTTGGCTCTCTGGCATTTGCAAGCTGTAACCGT	120
DB	4296	TTTGGGAACACGCGCGATCACTTCTTGGCTCTCTGGCATTTGCAAGCTGTAACCGT	4355
QY	121	ACCTTGGCTGCTCCCTCTTGGATTGAGTACAGCATCAAGCCCTCTTTTACCAACCTC	180
DB	4356	ACCTTGGCTGCTCCCTCTTGGATTGAGTACAGCATCAAGCCCTCTTTTACCAACCTC	4415
QY	181	CATGTGCTTACCAAGTACTTCAAGCTGGAGCCCTCCAGGCTTACATCGGGTATC	240
DB	4416	CATGTGCTTACCAAGTACTTCAAGCTGGAGCCCTCCAGGCTTACATCGGGTATC	4475
QY	241	AGCTTGGAGGATTTTCATGAGAGCTGGCACCACCCACTGCCCTCCCTGAGAGCGGGT	300
DB	4476	AGCTTGGAGGATTTTCATGAGAGCTGGCACCACCCACTGCCCTCCCTGAGAGCGGGT	4535
QY	301	GCATACCTGTTGAGGTGGCAGCCAGCGAAGCCAGATAGAGACCGTGCCCATGAAG	360
DB	4536	GCATACCTGTTGAGGTGGCAGCCAGCGAAGCCAGATAGAGACCGTGCCCATGAAG	4595
QY	361	GAGGAACCCCTTTGGCCCATCTTGGATACAGTTTCAATGAGTTTCAACAAAGTCGAG	420
DB	4596	GAGGAACCCCTTTGGCCCATCTTGGATACAGTTTCAATGAGTTTCAACAAAGTCGAG	4655
QY	421	CTTTTACAGCATTTCTCTGAGTCTTCTACAGAACAAATGAGCGAGATTTCT	480
DB	4656	CTTTTACAGCATTTCTCTGAGTCTTCTACAGAACAAATGAGCGAGATTTCT	4715
QY	481	CCAAAGAACATCCCGTCTGCTCCCTGCGAGAGCCCGCCAGCTTCCCGTCTCTAGAA	540
DB	4716	CCAAAGAACATCCCGTCTGCTCCCTGCGAGAGCCCGCCAGCTTCCCGTCTCTAGAA	4775
QY	541	GACACAGGCGCATACAGAGTACATGATGTTGGTACAGAAATGTTGAAGACGGGAG	600
DB	4776	GACACAGGCGCATACAGAGTACATGATGTTGGTACAGAAATGTTGAAGACGGGAG	4835
QY	601	GCCAGATTATGCCACCTTGTCCGCGCTTATGTGGCATTCATCTGGCAATGGCTCT	660
DB	4836	GCCAGATTATGCCACCTTGTCCGCGCTTATGTGGCATTCATCTGGCAATGGCTCT	4895

QY	661	GACTGGAAGAACGCTCTGTGCCATGCTGAAGGACGGGACTGAGGCTCGCACTTCATGGCC	720
DB	4896	GACTGGAAGAACGCTCTGTGCCATGCTGAAGGACGGGACTGAGGCTCGCACTTCATGGCC	4955
QY	721	TCTCCGAGTGTGTGGGTACAGCGCAGCAAGCGGCCCCCTCACGATGACTATGTGC	780
DB	4956	TCTCCGAGTGTGTGGGTACAGCGCAGCAAGCGGCCCCCTCACGATGACTATGTGC	5015
QY	781	CTGCTGACCTGAAGGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGGTGCGTGGATGC	840
DB	5016	CTGCTGACCTGAAGGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGGTGCGTGGATGC	5075
QY	841	CAGTCGCTCTAGCTTGTCTACTGATTCGAGAGTATGTGCCTGAGCTCCAAAGCTCTTC	900
DB	5076	CAGTCGCTCTAGCTTGTCTACTGATTCGAGAGTATGTGCCTGAGCTCCAAAGCTCTTC	5135
QY	901	AAAGGGAAGGTGAAGGTGTGAGCCTGAAGCCTGAGGTGGCCAGGTGCACTGTACATC	960
DB	5136	AAAGGGAAGGTGAAGGTGTGAGCCTGAAGCCTGAGGTGGCCAGGTGCACTGTACATC	5195
QY	961	CTGCGCCCAAGCGGACCTTTATTGGCAACTGTGTCTCTCTTCACTGCCCTTTGTGAAG	1020
DB	5196	CTGCGCCCAAGCGGACCTTTATTGGCAACTGTGTCTCTCTTCACTGCCCTTTGTGAAG	5255
QY	1021	CGGAGCGGACCTCCAGGGAGGCGCTCTTCTTTCGGCATGCAAGGCCCTCTAAG	1080
DB	5256	CGGAGCGGACCTCCAGGGAGGCGCTCTTCTTTCGGCATGCAAGGCCCTCTAAG	5315
QY	1081	CTGCGGACGAGTCTTGATTTCTGCGCGAGCACCAAGCCCTGTGATCTCTGGAGGACCA	1140
DB	5316	CTGCGGACGAGTCTTGATTTCTGCGCGAGCACCAAGCCCTGTGATCTCTGGAGGACCA	5375
QY	1141	AGTCTGAGCTGCTCTTCCAGCCAGGCTGCGACGAGGTGCTCCGGGATTCGAACT	1200
DB	5376	AGTCTGAGCTGCTCTTCCAGCCAGGCTGCGACGAGGTGCTCCGGGATTCGAACT	5435
QY	1201	CCTCTTCTCACTGCAAGATGGAAGAGTGGCCAGGAGCCCTCAAGGAGGAGAGCG	1260
DB	5436	CCTCTTCTCACTGCAAGATGGAAGAGTGGCCAGGAGCCCTCAAGGAGGAGAGCG	5495
QY	1261	TCCATATCCAGGCAATAGGACTTGCAGGTTCTTAGGAGGAGGAGCATCTCCCATCGAC	1320
DB	5496	TCCATATCCAGGCAATAGGACTTGCAGGTTCTTAGGAGGAGGAGCATCTCCCATCGAC	5555
QY	1321	GTGCTTCTGCTCTTCTGGGAAATTTCTCAGCTGGCAAGCAGTCCAGCTCCGCTTCT	1380
DB	5556	GTGCTTCTGCTCTTCTGGGAAATTTCTCAGCTGGCAAGCAGTCCAGCTCCGCTTCT	5615
QY	1381	GGTCCACTCTGCTGAGCAGCTGGGATGCTGAATCTTTCAGAGAGATTTTTTATAGA	1440
DB	5616	GGTCCACTCTGCTGAGCAGCTGGGATGCTGAATCTTTCAGAGAGATTTTTTATAGA	5675
QY	1441	GAGATTTCTAATTTTGTATACAGGTCATGACTATCTAGAACTCTCTGGTTTTTGA	1500
DB	5676	GAGATTTCTAATTTTGTATACAGGTCATGACTATCTAGAACTCTCTGGTTTTTGA	5735
QY	1501	AAATCAATGAATTC	1514
DB	5736	AAATCAATGAATTC	5749

RESULT 3  
US-10-301-822-168  
; Sequence 168, Application US/10301822  
; Publication No. US20030148410A1  
; GENERAL INFORMATION:  
; APPLICANT: Millennium Pharmaceuticals, Inc.  
; APPLICANT: Berger, Allison  
; APPLICANT: Guillemette, Tracy L.  
; APPLICANT: Kamatkar, Shubhangi  
; APPLICANT: Schlegel, Robert  
; APPLICANT: Monahan, John E.  
; APPLICANT: Thibodeau, Stephen N.

APPLICANT: Burgart, Lawrence J.  
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND  
METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
THERAPY OF COLON CANCER  
FILE REFERENCE: MP01-029P2RNM  
CURRENT APPLICATION NUMBER: US/10/301,822  
CURRENT FILING DATE: 2002-11-21  
PRIORITY APPLICATION NUMBER: US 60/339,971  
PRIORITY FILING DATE: 2001-12-10  
PRIORITY APPLICATION NUMBER: US 60/361,978  
PRIORITY FILING DATE: 2002-03-05  
PRIORITY APPLICATION NUMBER: US 60/381,988  
PRIORITY FILING DATE: 2002-05-20  
NUMBER OF SEQ ID NOS: 228  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 168  
LENGTH: 5218  
TYPE: DNA  
ORGANISM: Homo Sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (50)... (1216)  
US-10-301-822-168

Query Match 99.8%; Score 1510.8; DB 16; Length 5218;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1512; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
1 ATGCCCGCGGCTCTGGGACCGGCGGTTACCTGCTCTACTGCCCCCTGCGATGGGCGC 60  
119 ATGCCCTGGGCTCTGGGACCGGCGGTTACCTGCTCTACTGCCCCCTGCGATGGGCGC 178  
61 TTGGGACACGCGGATCACTCTTGGGCTCTCTGGCATTTGCAAGCTGTAACCGT 120  
179 TTGGGACACGCGGATCACTCTTGGGCTCTCTGGCATTTGCAAGCTGTAACCGT 238  
121 ACCTGGCTGCTCTCTGGATTTGATGATCAGGATCAGAGCTCTCTTACCACTC 180  
239 ACCTGGCTGCTCTCTGGATTTGATGATCAGGATCAGAGCTCTCTTACCACTC 238  
181 CATGTCCTACAGAGTATTCAGTGTGAGGCCCCCTCAGGCTTACCATCGGTCATC 240  
299 CATGTCCTACAGAGTATTCAGTGTGAGGCCCCCTCAGGCTTACCATCGGTCATC 358  
241 AGCTTGGAGGATTCATGGAGAGCTGGACCCACCCACTGGCCCCCTCAGAGCGGGTG 300  
359 AGCTTGGAGGATTCATGGAGAGCTGGACCCACCCACTGGCCCCCTCAGAGCGGGTG 418  
301 GCATATCTCTTGGAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 360  
419 GCATATCTCTTGGAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 478  
361 GAAGGAACCCCTTGGGCTCTGGGATCAGTTTCTATGTGAGTTTCAACAAGTCGGAG 420  
479 GAAGGAACCCCTTGGGCTCTGGGATCAGTTTCTATGTGAGTTTCAACAAGTCGGAG 538  
421 CTTTTCAGGCAATTCCTTCACTGCTCTGATGATGATGATGATGATGATGATGATGAT 480  
539 CTTTTCAGGCAATTCCTTCACTGCTCTGATGATGATGATGATGATGATGATGATGAT 598  
481 CCAAGGAACATCCGCTGCTGGGCTGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 540  
599 CCAAGGAACATCCGCTGCTGGGCTGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 658  
541 GAACACAGGCACTACAGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 600  
659 GAACACAGGCACTACAGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 718  
601 GCCAGATTCATGCCACCTTGTCCGCGCTATGTGGGCAATTCATCTGGGCAATTCGCTCT 660  
719 GCCAGATTCATGCCACCTTGTCCGCGCTATGTGGGCAATTCATCTGGGCAATTCGCTCT 778  
661 GACTGGAAGACGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720

779 GACTGGAAGACGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 838  
721 TCTCCGCAAGTGTGTGGGCTACAGCCGAGCAGACAGCGGCCCTCCATGATGATGTC 780  
839 TCTCCGCAAGTGTGTGGGCTACAGCCGAGCAGACAGCGGCCCTCCATGATGATGTC 898  
781 CTGCTCTGACCTGAAGGAGATCAGAGGCGTGTGAAGCTCTGGGTGAGTCTGCTGATGCC 840  
899 CTGCTCTGACCTGAAGGAGATCAGAGGCGTGTGAAGCTCTGGGTGAGTCTGCTGATGCC 958  
841 CAGTCGGTCTACGTTGCTACTGATTTCCGAGAGTTATGTGCTGAGCTCCAAAGCTCTTC 900  
959 CAGTCGGTCTACGTTGCTACTGATTTCCGAGAGTTATGTGCTGAGCTCCAAAGCTCTTC 1018  
901 AAAGGAAAGTGAAGTGTGAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGA 960  
1019 AAAGGAAAGTGAAGTGTGAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGA 1078  
961 CTCGCCAAGCCGACCACTTTATTTGGCAACTGTGTCTCTCTCTCTCTCTCTCTCTCTCT 1020  
1079 CTCGCCAAGCCGACCACTTTATTTGGCAACTGTGTCTCTCTCTCTCTCTCTCTCTCTCT 1138  
1021 CGGGAGCGGACCTCCAGGGGAGGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1080  
1139 CGGGAGCGGACCTCCAGGGGAGGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1198  
1081 CTGCGGAGCGAGTTCTGATTTCTGGCGGAGCACAGACCTCTCTCTCTCTCTCTCTCTCTCT 1140  
1199 CTGCGGAGCGAGTTCTGATTTCTGGCGGAGCACAGACCTCTCTCTCTCTCTCTCTCTCTCT 1258  
1141 AGTCTGAGCTGCTCTCTCAAGCCAGGCTTGGCAGCAGAGGTGCTCGGGATTGCAAACT 1200  
1259 AGTCTGAGCTGCTCTCTCAAGCCAGGCTTGGCAGCAGAGGTGCTCGGGATTGCAAACT 1318  
1201 CCTCTTCTACCTGCGCAAGATGGAGAGTGGCCAGGACCCCTCAAGGAGGAGAGCGC 1260  
1319 CCTCTTCTACCTGCGCAAGATGGAGAGTGGCCAGGACCCCTCAAGGAGGAGAGCGC 1378  
1261 TCATATCCAGGGCATAGGACTTGGCAGGTTCTTAGAGCAGAGGATCTCCCATGCGAC 1320  
1379 TCATATCCAGGGCATAGGACTTGGCAGGTTCTTAGAGCAGAGGATCTCCCATGCGAC 1438  
1321 GTGCTTTCTGCTCTCTGGGAATTTCTCACTGGCAAGAGTCCAGGCTCCGCTCTTCT 1380  
1439 GTGCTTTCTGCTCTCTGGGAATTTCTCACTGGCAAGAGTCCAGGCTCCGCTCTTCT 1498  
1381 GGTCCACTCTCTCTGAGCAGCCTGGGATGCTGAACCTTTTCAAGAGATTTTTTATAGA 1440  
1499 GGTCCACTCTCTCTGAGCAGCCTGGGATGCTGAACCTTTTCAAGAGATTTTTTATAGA 1558  
1441 GAGATTTCTATAATTTTATACAAGGTATGATGATGATGATGATGATGATGATGATGATGAT 1500  
1559 GAGATTTCTATAATTTTATACAAGGTATGATGATGATGATGATGATGATGATGATGATGAT 1618  
1501 AAATCATTTGAATTC 1514  
1619 AAATCATTTGAATTC 1632

RESULT 4  
US-10-956-157-4752  
; Sequence 4752, Application US/10956157  
; Publication No. US20050118625A1  
; GENERAL INFORMATION:  
; APPLICANT: Wyeth  
; APPLICANT: Mounts, William  
; TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION ASSOCIATED WITH  
; FILE REFERENCE: 031896-043000 (AM 101081)  
; CURRENT APPLICATION NUMBER: US/10/956,157  
; CURRENT FILING DATE: 2004-10-04  
; NUMBER OF SEQ ID NOS: 319805



; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 4752  
; LENGTH: 5218  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-956-157-4752

Query Match 99.8%; Score 1510.8; DB 22; Length 5218;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1512; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGCGCGGGCTCTCGGACCCGCGGTACCTGCTCTACTGCCCCCTGATGCGGGCGC 60  
DB 119 ATGCGCGGGCTCTCGGACCCGCGGTACCTGCTCTACTGCCCCCTGATGCGGGCGC 178  
QY 61 TTGCGGAACAGCGCGATCACTTCTTGGCTCTCTGGCATTTGCAAACTGCTAAACCGT 120  
DB 179 TTGCGGAACAGCGCGATCACTTCTTGGCTCTCTGGCATTTGCAAACTGCTAAACCGT 238  
QY 121 ACCTTGGCTGCTCCCTCTTGGATTGAGTACAGCATCAAGCCCTCTTTTCAACCAACCTC 180  
DB 239 ACCTTGGCTGCTCCCTCTTGGATTGAGTACAGCATCAAGCCCTCTTTTCAACCAACCTC 298  
QY 181 CATGTGCTTACAGAACTTCAAGCTGGAGCCCTCCAGGCTTACATCGGGTCATC 240  
DB 299 CATGTGCTTACAGAACTTCAAGCTGGAGCCCTCCAGGCTTACATCGGGTCATC 358  
QY 241 AGCTTGGAGGATTTTCATGAGAACTGCGACCCACCCACTGCGCCCTGAGAGCGGGTG 300  
DB 359 AGCTTGGAGGATTTTCATGAGAACTGCGACCCACCCACTGCGCCCTGAGAGCGGGTG 418  
QY 301 GCATCTGCTTTGAGTGGCGACCCAGCGAAGCCAGATAGAGACCGTGCCTCATGAAG 360  
DB 419 GCATCTGCTTTGAGTGGCGACCCAGCGAAGCCAGATAGAGACCGTGCCTCATGAAG 478  
QY 361 GAAGAAACCCCTTTGGCCCATTTCTGGGATCAGTTTTCATGTAGTTTCAAAAGTCGGAG 420  
DB 479 GAAGAAACCCCTTTGGCCCATTTCTGGGATCAGTTTTCATGTAGTTTCAAAAGTCGGAG 538  
QY 421 CTTTTTACAGCATTTCTCTCAGTCTTCTTACAGAGCAATGAGCCAGATTTTCT 480  
DB 539 CTTTTTACAGCATTTCTCTCAGTCTTCTTACAGAGCAATGAGCCAGATTTTCT 598  
QY 481 CCAAGGAACATCCGGTCTGCTGCGCCCTGCGAGAGCCCGAGTCTCCCGTCTCTAGAA 540  
DB 599 CCAAGGAACATCCGGTCTGCTGCGCCCTGCGAGAGCCCGAGTCTCCCGTCTCTAGAG 658  
QY 541 GAACACAGGCCACTACAGAACTACATGTTGTTGAGTGGTGAAGACGGGAGAG 600  
DB 659 GAACACAGGCCACTACAGAACTACATGTTGTTGAGTGGTGAAGACGGGAGAG 718  
QY 601 GCCCAGATTATGCCACCTTGTCCGCGCCCTATGTGGGCATTCATCTCGGCATTTGGCTCT 660  
DB 719 GCCCAGATTATGCCACCTTGTCCGCGCCCTATGTGGGCATTCATCTCGGCATTTGGCTCT 778  
QY 661 GACTGGAAGAACCGCTGTGCTCATGTGGAAGACGGGACTGCGAGGCTGCCTCATGGCC 720  
DB 779 GACTGGAAGAACCGCTGTGCTCATGTGGAAGACGGGACTGCGAGGCTGCCTCATGGCC 838  
QY 721 TCTCCGAGTGTGCGGTACAGCGCGACAGCGGCCCTTACGATGACTATGTGC 780  
DB 839 TCTCCGAGTGTGCGGTACAGCGCGACAGCGGCCCTTACGATGACTATGTGC 898  
QY 781 CTGCTGACCTGAAGAGATCCAGAGGCTGTGAAGCTCTGGTGAGTCTCGTGGATGCC 840  
DB 899 CTGCTGACCTGAAGAGATCCAGAGGCTGTGAAGCTCTGGTGAGTCTCGTGGATGCC 958  
QY 841 CAGTCGGTCTAGTGTGCTACTGATTCGAGAGTATGTGCTGTGAGCTCCAAACAGCTCTTC 900  
DB 959 CAGTCGGTCTAGTGTGCTACTGATTCGAGAGTATGTGCTGTGAGCTCCAAACAGCTCTTC 1018  
QY 901 AAAGGAAGGTGAAGGTGGTGAAGCTGAGGCTGAGGTGGCCAGGTCGACTGTACATC 960

DB 1019 AAAGGAAGGTGAAGGTGGTGAAGCTGAGGTGGCCAGGTCGACCTGTACATC 1078  
QY 961 CTGCGCAAGCGGACCACTTTATTGGCAACTGTGCTCTCTCTTCACTGCGCTTGTGAAG 1020  
DB 1079 CTGCGCAAGCGGACCACTTTATTGGCAACTGTGCTCTCTCTTCACTGCGCTTGTGAAG 1138  
QY 1021 CCGGAGCGGACCTCCAGGGGAGGCGCTTCTTTCTTTCGCGCATGGACAGGCCCTTAAG 1080  
DB 1139 CCGGAGCGGACCTCCAGGGGAGGCGCTTCTTTCTTTCGCGCATGGACAGGCCCTTAAG 1198  
QY 1081 CTGCGGAGCAGTGTCTGATTTCTGCGCGGAGCACAGACCTCTGATCTCGAGGAGACGAG 1140  
DB 1199 CTGCGGAGCAGTGTCTGATTTCTGCGCGGAGCACAGACCTCTGATCTCGAGGAGACGAG 1258  
QY 1141 AGCTGAGCTGCTCTTCCAGCGGCTGCGCAGCAGAGGTGCTCCGGGATTTGCAAACT 1200  
DB 1259 AGCTGAGCTGCTCTTCCAGCGGCTGCGCAGCAGAGGTGCTCCGGGATTTGCAAACT 1318  
QY 1201 CCTCTTCTCACTGCGCAAGATGGAGAGAGTGGCCAGGAGCCCTTCAAGGAGGAGAGCGC 1260  
DB 1319 CCTCTTCTCACTGCGCAAGATGGAGAGAGTGGCCAGGAGCCCTTCAAGGAGGAGAGCGC 1378  
QY 1261 TCCATATCCAGGGCATAGGACTTTCAGGTTCTTAGGAGCAGGAGCATCTCCCATCGCAC 1320  
DB 1379 TCCATATCCAGGGCATAGGACTTTCAGGTTCTTAGGAGCAGGAGCATCTCCCATCGCAC 1438  
QY 1321 GTGCTTTCTGCTCTTCTGGGAAATTTCTCACACTGGCAAGCAGTCCAGCTCCGCTTCT 1380  
DB 1439 GTGCTTTCTGCTCTTCTGGGAAATTTCTCACACTGGCAAGCAGTCCAGCTCCGCTTCT 1498  
QY 1381 GGTCCACTCTCTCTGAGCAGCTGGGATGCTGAACTCTTTCAGAGAGATTTTTTATAGA 1440  
DB 1499 GGTCCACTCTCTCTGAGCAGCTGGGATGCTGAACTCTTTCAGAGAGATTTTTTATAGA 1558  
QY 1441 GAGATTTCTATAATTTTGTATACAAAGTTCATGCTATCTCTAGAACTCTCTGTGGTTTTTGA 1500  
DB 1559 GAGATTTCTATAATTTTGTATACAAAGTTCATGCTATCTCTAGAACTCTCTGTGGTTTTTGA 1618  
QY 1501 AAATCATTTGAATTC 1514  
DB 1619 AAATCATTTGAATTC 1632

RESULT 5

US-09-774-954-7  
; Sequence 7, Application US/09774954  
; Publication No. US20040241645A1  
; GENERAL INFORMATION:  
; APPLICANT: Yang Wang, Michael W. Spellman  
; TITLE OF INVENTION: O-Pucosyltransferase  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genentech, Inc.  
; STREET: 1 DNA Way  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WinPatIn (Genentech)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/774,954  
; FILING DATE: 30-Jan-2001  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/978,741  
; FILING DATE: 26-NOV-1997  
; APPLICATION NUMBER: 08/792,498  
; FILING DATE: 31-JAN-1997  
; ATTORNEY/AGENT INFORMATION:







[illegible]



Qy	197	AGTACTTCAAGCTGGAGCCCTCAGGGCTTACCATTACATCGGGTCATCAGCTTGGAGGATTTC	256
Db	254	CATATTTGAAGTGGAGCCCTGAAGGAATACCATCGCGTCATCACCATGGCAGATTTC	313
Qy	257	TGGAGAAGCTGGCACCCACCCACTGGCCCCCTGTAGAGAGCGGGTGGCATACTGCTTTGAGG	316
Db	314	TGTGGCACCTGGCCGACACATTTGGCCAGAAATCGGACGAGTGTTCATTTTGTTCACAAGG	373
Qy	317	-----TGGCAGCCCGACGAGCCAGATAAGAAAGACGTGCCCCATGAAGG	361
Db	374	AACGATATAGCTTTCAGCAGGAGAGAACGATCCAGACAAGCCCAATTGCCACGCCAAGG	433
Qy	362	AAGGAACCCCTTTGGGCCATTCTGGGATCAGTTTTCATGTGAGTTTCAACAAGTCGGAGC	421
Db	434	ATGCAATTCCTTTTGGTCCCTTTTGGGACACTTTTTCACATAGACTTTTGTGCGGTGACAGT	493
Qy	422	TTTTTACAGGATTTCTCTTCAGTGTCTCCTACAG-----AGAACAAATGGAGCCAGA	472
Db	494	TCTATGGCCACTTTCATTTTGATGTGATCATAGCAACGAGGCTGCCAAGTGGCAGACCA	553
Qy	473	GATTTTCTCCAAAGGAACATCCGGTGTCTGCCCTGCCAGGAGCCCCAGCCAGTTCGCCG	532
Db	554	AATATCTGCAGATCATATCCCGTACTCCGGTTTACCCGAGCTCCCGGTAGTTTTCCTG	613
Qy	533	TCCTAGAAGAACACAGGCCACTACAGAAAGTACATGGTATGTCTCAGACGAAATGGTGAAGA	592
Db	614	TTCAGCTAGAGAACTGCAAGCTGACGCGCTACTTTGCAGTGAGTCAACGGTATAGGAG	673
Qy	593	CGGAGAGGCCAGATTATGCCCACTGTGCCGC---CCTATGTGGGCATTCATCTGC	649
Db	674	CATCTAAGGATTTTCATCCGAGACAGTTGCCCTCGGGGTGCCCTTTTGGGCATTCATCTGC	733
Qy	650	GCATTGGCTCTGACTGGAGNAAGCCCTGTGCCATGCTGAAGGACGG-----GACTG	700
Db	734	GCAACGTATCGATTGGGTGAGAGCCCTGTGAGCACGTCGAGGATAGCCAGCATCTGTTTG	793
Qy	701	CAGGCTCGCACTTCATGGCCTCTCCGCAGTGTGTGGCTTACAGCGGACGACAGCGGCC	760
Db	794	CCTCGCCGCACTGTGGGTATAAAATGAACTGGTGTCACTCTACCCGAGCTCTGCA	853
Qy	761	CCCTCAGATGACTATATGCTCGCTGCCTGACCTGAAGGAGATCCAGAGGCTGTGAAGTCT	820
Db	854	TGCCCTCCAAGGAGCGATCATCCGCGACTAAAGAGAACCATTAAGAACGTGCGCCAAA	913
Qy	821	GGGTGAGTCTGTGATGCCAGTTCGGTCTACGTTGCTATCTGATTCGAGAGTTATGTGC	880
Db	914	CTCAGCCGACAAAGAAATCAAAATCAGTTTTCGTGGGCTCAGACTCCAATCACATGATTG	973
Qy	881	CTGAGCTCCAACAGCTCTTCAAGGGAGGTGAAGTGGTTCAGCCTGAAGCCTCAGGTGG	940
Db	974	GTGAACTAACACGGCCCTTAGTCGATGGGCATCAGTGTGCACAAGTGCCTGGAGGATG	1033
Qy	941	---CCCAGGTGCACTGTGATCATCTCGGCCAAGCGGACCACTTTATTTGGCAACTGTGTCT	997
Db	1034	ATCCTTACCTGGACTTGGCCATTTCTCGGACAGTCAACCACTTTATCGGCNACTGTATAT	1093
Qy	998	CCTCCTTACTGCTTTGTGAAGCGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTCT	1057
Db	1094	CCTCTTACTCGGCATTCGAAAAAAGGAAACGAGATGTGCAGGTTTTCATCGTACTTCT	1153
Qy	1058	TCGGCATGGACAGG	1071
Db	1154	GGGGATTTCCCAAG	1167

RESULT 14  
US-11-097-143-18160/c  
; Sequence 18160, Application US/11097143  
; Publication No. US20050208558A1  
; GENERAL INFORMATION:  
; APPLICANT: Venter, J. Craig  
; APPLICANT: et al.



```
Qy 625 CGG---CCCTATGTGGCATTTCATCTGGCGATTGGCTCTGACTGGAAGACGCCCTGTGCC 681
Db 1504 CGGGGTGCTTTTGTGGGCATTTCATCTGGCGCAACGGTATCGATTGGGTGAGAGCCTGTGAG 1445
Qy 682 ATGCTGGAAGACGG-----GACTGCAAGCTCGCAGCTTCATGGCCCTCTCCGAGTGT 732
Db 1444 CAGCTCAAGGATAGCCAGCATCTGTTTGCCTCGCGCAGTGTCTGGGCTATAAAAATGAA 1385
Qy 733 GTGGGCTACAGCGCAGCAGCGCGCCCTCCAGATGACTATGTGCTGCTGCTGACCTG 792
Db 1384 CGTGTGCACTTACCCGGAGCTTCGATGCCCTCCAGGAGCGCATCATCGCCAGCTA 1325
Qy 793 AAGGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGTTCGATGCCAGTCCGCTCTAC 852
Db 1324 AAGAAACCATTAAGACGTGCGCAACTCAGCGGACACGAAATCAATCAGTTTC 1265
Qy 853 GTTGCTACTGATTCGAGAGTTATGTGCTGAGCTCCAGAGCTCTTCAAGGAGGTG 912
Db 1264 GTGGCGTCAGACTCCAATCATGATTTGGTGAACCTAAACAGCCCTTAGTCGCAATGGC 1205
Qy 913 AAGGTGTGAGCCTGAAGCCTGAGTGG---CCAGGTGCACTGTACATCTCTCGCCAA 969
Db 1204 ATCAGTGTGCAAGCTGCGGAGGATGATCTTACCTGGAGCTTGGCCATCTCGACAG 1145
Qy 970 GCCGACCACTTTATTTGGCACTGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1029
Db 1144 TCGAACCACTTTATCGGCAACTGTATATCTCTTACTCGGCATTCGAAAAAAGGAACGA 1085
Qy 1030 GACCTCCAGGGAGGCGCTCTCTTTCTTGGGCATGGACAGG 1071
Db 1084 GATGTGACGCTTTTCCATCGTACTTCTGGGATTCGCCAAG 1043
```

RESULT 15

```
US-09-783-590-11501
; Sequence 11501, Application US/09783590
; Patent No. US20020110850A1
; GENERAL INFORMATION:
; APPLICANT: Dillion, Patrick J.
; APPLICANT: Haseltine, William A.
; APPLICANT: Li, Haodong
; APPLICANT: Rosen, Craig A.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
; FILE REFERENCE: PO-16.2C1
; CURRENT APPLICATION NUMBER: US/09/783,590
; CURRENT FILING DATE: 2000-02-15
; PRIOR APPLICATION NUMBER: 08/420,856
; PRIOR FILING DATE: 1995-04-12
; PRIOR APPLICATION NUMBER: 08/346,731
; PRIOR FILING DATE: 1994-11-21
; NUMBER OF SEQ ID NOS: 12485
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11501
; LENGTH: 402
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (9)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (19)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (22)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (46)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (272)
```

```
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (301)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (320)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (333)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (342)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (374)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (386)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (390)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (400)
; OTHER INFORMATION: n equals a,t,g, or c
; US-09-783-590-11501

Query Match 7.6%; Score 114.6; DB 9; Length 402;
Best Local Similarity 83.9%; Pred. No. 5.7e-26;
Matches 177; Conservative 0; Mismatches 23; Indels 11; Gaps 4;

Qy 178 CTCCATGTGCTCTACCAGAAAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATGGGTC 237
Db 161 CTCCATGTGCTCTACCAGAAAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATGGGTC 220
Qy 238 ATCAGCTT-GGAGGATTTTCATGGAGAAGCTGGCACCCACCCACCTGGCCCCCT---GAGAA 293
Db 221 ATCAGCTTGGGAGGATTTTCATGGAGAAGCTGGCACCCACCCACCTGGCCCCCTNTGAGAAGC 280
Qy 294 GCGGTGGCATACTGCTTTGA---GGTGGCAGCCAGCGAAGCCAGAT-----AAGAAGA 346
Db 281 GCGGTGGCATAATGCTTTTNNAGTGGGCGAGCCCGAGCCGANGCCAGATTAAAGNAAGGACG 340
Qy 347 CGTGGCCCATGAAGGAAGAACCCCTTTGG 377
Db 341 TNGACCCCATGAAGGAAGAACCCCTTTTG 371
```

Search completed: November 9, 2005, 15:51:13  
Job time : 1357.3 secs





109 ATGCTCGGGCTCTCTGGACCGCGCGGTACCTGCTCTACTGCCCCCTGCATGGGGCGC 168  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
841  
842  
843  
844  
845  
846  
847  
848  
849  
850  
851  
852  
853  
854  
855  
856  
857  
858  
859  
860  
861  
862  
863  
864  
865  
866  
867  
868  
869  
870  
871  
872  
873  
874  
875  
876  
877  
878  
879  
880  
881  
882  
883  
884  
885  
886  
887  
888  
889  
890  
891  
892  
893  
894  
895  
896  
897  
898  
899  
900  
901  
902  
903  
904  
905  
906  
907  
908  
909  
910  
911  
912  
913  
914  
915  
916  
917  
918  
919  
920  
921  
922  
923  
924  
925  
926  
927  
928  
929  
930  
931  
932  
933  
934  
935  
936  
937  
938  
939  
940  
941  
942  
943  
944  
945  
946  
947  
948  
949  
950  
951  
952  
953  
954  
955  
956  
957  
958  
959  
960  
961  
962  
963  
964  
965  
966  
967  
968  
969  
970  
971  
972  
973  
974  
975  
976  
977  
978  
979  
980  
981  
982  
983  
984  
985  
986  
987  
988  
989  
990  
991  
992  
993  
994  
995  
996  
997  
998  
999  
1000

QY 1141 AGTCTGAGCTGGTCTTCCAGCCAGGCTTGGCAGAGGCTCTCCGGGATTGCAAACT 1200  
DB 1249 AGTCTGAGCTGGTCTTCCAGCCAGGCTTGGCAGAGGCTCTCCGGGATTGCAAACT 1308  
QY 1201 CCTCTTCTCACCCTGCGCAAGAGAGTGCAGGACCCCTCAAGAGGGAGAGCGC 1260  
DB 1309 CCTCTTCTCACCCTGCGCAAGAGAGTGCAGGACCCCTCAAGAGGGAGAGCGC 1368  
QY 1261 TCATATCCAGGGCATAGGACTTGCAGTTCCTAGGAGCAGGAGCATCTCCATCGCAC 1320  
DB 1369 TCATATCCAGGGCATAGGACTTGCAGTTCCTAGGAGCAGGAGCATCTCCATCGCAC 1428  
QY 1321 GTCCTTCTGCTCTTCTGGGAATTTCTCAGCTGGGAAAAGCAGTCCAGCTCCGTTCT 1380  
DB 1429 GTCCTTCTGCTCTTCTGGGAATTTCTCAGCTGGGAAAAGCAGTCCAGCTCCGTTCT 1488  
QY 1381 GTCCTTCTGCTCTTCTGGGAATTTCTCAGCTGGGAAAAGCAGTTCAGAGAGATTTTATAGA 1440  
DB 1489 GTCCTTCTGCTCTTCTGGGAATTTCTCAGCTGGGAAAAGCAGTTCAGAGAGATTTTATAGA 1548  
QY 1441 GAGATTT 1447  
DB 1549 GAGATTT 1555

RESULT 2  
BX353274 1141 bp mRNA linear EST 08-APR-2004  
LOCUS BX353274 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens  
DEFINITION CDNA Clone CS0DC003JY11 5-PRIME, mRNA sequence.  
ACCESSION BX353274  
VERSION BX353274.2 GI:46290116  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1141)  
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
Full-length cDNA libraries and normalization  
Unpublished (2001)  
On May 5, 2003 this sequence version replaced gi:30371746.  
COMMENT Contact: Genoscope  
Genoscope - Centre National de Sequencage  
2 rue Gaston Crémieux, CP 5706 - 91057 EVRY cedex - FRANCE  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
end enriched, double-strand cDNA was digested with Not I and cloned  
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
was normalized. Library was constructed by Life Technologies, a  
division of Invitrogen. This sequence belongs to sequence cluster  
9980.f  
For more information about this cluster, see  
http://www.genoscope.cns.fr/cdna?s=CS0DC003CE06QPl&c=9980.f.

## FEATURES

source  
1..1141  
location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clones="CS0DC003JY11"  
/tissue\_type="NEUROBLASTOMA COT 25-NORMALIZED"  
/clone\_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo(dT)  
primer. Five prime end enriched, double-strand cDNA was  
digested with Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
sites of the pCMVSPORT 6 vector. Library was normalized."

## ORIGIN

Query Match 59.0%; Score 892.8; DB 5; Length 1141;  
Best Local Similarity 92.8%; Pred. No. 2.6e-236;  
Matches 952; Conservative 32; Mismatches 35; Indels 7; Gaps 5;



Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Please visit our web site for further details. URL: <http://genome.gsc.riken.jp/> URL: <http://fantom.gsc.riken.jp/>

FEATURES		Location/Qualifiers	
source	1..3611	organism="Mus musculus"	
		mol_type="mRNA"	
		strain="C57BL/6J"	
		db_xref="FANTOM_DB:B930076G19"	
		db_xref="taxon:10090"	
		clone="B930076G19"	
		tissue_type="cerebellum"	
		/clone_lib="RIKEN full-length enriched mouse cDNA library"	
misc_feature	11..1194	/note="protein O-fucosyltransferase 1 (MGDI:2153207, GB NM_080463, evidence: BLASTN, 99%, match=1290) putative"	
ORIGIN			
Query Match 58.9%; Score 892; DB 3; Length 3611;			
Best Local Similarity 83.3%; Pred. No. 6.3e-236;			
Matches 1051; Conservative 0; Mismatches 180; Indels 31; Gaps 2;			
QY	5	CCGCGGGCTCTCGGAGCCCGCGGTTACTGCTCTACTGCTCCCTGCATGGGGCGCTTTG	64
DB	101	CAGCGGGCTCTCGGAGCCTCGCGCGTTACTGCTCTACTGCTCCCTGCATGGGGCGCTTTG	160
QY	65	GGAAACAGAGCCGATCACTCTTCTGGGCTCTCTGGCATTTGGCAAGCTCTCTAAACCGGTACT	124
DB	161	GGAAACAGAGCTGATCACTCTTCTGGGCTCTCTGGCATTTGGCAAGCTCTCTGAAACCGCACT	220
QY	125	TGGCTGTCCCTCTTGGGATTCAGTATACAGAGCATCACAAAGCCTCTCTTCAACACCTCCATG	184
DB	221	TGGCTGTACTCTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	280
QY	185	TGCTCTACAGAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGGTCAATCAGCT	244
DB	281	TGCTCTACAGAGTACTTCAAGCTGGAGCCCTCTCCAGGCTTACCATCGGGTGTTCAGGCC	340
QY	245	TGAGGAGTTTCATGGAGAGCTGGACCCACCCCTGCGCCCTCTGAGAGCGGGTGGCAT	304
DB	341	TGAGGAGTTTCATGGAGAGTTCTGGACCCCTCCCATGGCCCTCTGAGAGGGAGTGGCAT	400
QY	305	ACTGCTTTGAGTGCGAGCCAGCGAGCCAGATGAAGACGCTGCCCATCAAGAGGAG	364
DB	401	ACTGCTTTGAGTGCGAGCCAGCGAGTCTCTGATGAAGACATGTCCCATGAAGAGAG	460
QY	365	GAACCCCTTTGGCCCATCTCTGGGATCAGTTTCAATGTGAGTTTCAACAGTCGAGCTTT	424
DB	461	GAATCTCTTTTGGGCACTTCTGGGACAGTTTCAATGTGAGTTTCAATGAGTCAAGACTGT	520
QY	425	TTACAGGCAATTCCTTTCAGTGTCTCTACAGAGAAACAATGGAGCCAGAGATTTTCTCNA	484
DB	521	TCACAGGCAATTTCTTTCAGCGCTCTCTACAAAGAAACAATGGACCCAGAGATTTTCTGCA	580
QY	485	AGGAACATCCGGTGTGCTTGGCCCTGCGAGGAGCCCGAGCCAGTTCCCGTCTCTAAGAAGAA	544
DB	581	AAGAGCATCTCTGTCTCGCACTCTCTGGGGCCCGAGCAGTTCTCTGTCTGAGAGAG	640
QY	545	ACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGCGGAGAGGCC	604
DB	641	ACAGGGAGCTCCAGAAGTACATGGTGTGGTCAGATGAGATGGTGAAGCGGAGAGGCC	700
QY	605	AGATTTCATGCCCACTTTGTCGGGCCCTATGTGGGCATTCATCTGGCGCATTTGGCTCTGACT	664
DB	701	TGATCAGTGCCCACTCTGTCGGGCCCTATGTGGGCATTCATCTGGCGCATTTGGCTCTGACT	760
QY	665	GGAGAGAGCGCTCTGGCATGCTGAAGAGCGGAGCTGCGAGCTCGCACTTCATGAGCCCTCTC	724

Db	761	GGAGAGATGCTTGCCATGCTGGAAGGATGAACTGAGGGTCACTTCTGCTTCCC	820
Qy	725	CGAGTGTGTGGGCTACAGCCGACAGCGGGCCCCCTCAAGATGATGTGCTGTC	784
Db	821	CTAGTGTGTGGGCTATAGCCGACAGCGAGCCCTCTCACCATGACATGTGCTTCC	880
Qy	785	CTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCTGCTGATGCC	844
Db	881	CTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCTGCTGATGCC	940
Qy	845	CGCTCTAGCTTGTCTACTGATTTCCGAGAGTATGCTGCTGCTGCTGCTGCTGCTG	904
Db	941	CGCTCTAGCTTGTCTACTGATTTCCGAGAGTATGCTGCTGCTGCTGCTGCTGCTG	1000
Qy	905	GGAGGTGAAGGTGCTGAGCCTGAGCCTGAGCCTGAGGTGGCCAGGTGCTGATCTTC	964
Db	1001	ACAGGTGAAGGTGCTGAGCCTGAGCCTGAGCCTGAGGTGGCCAGGTGCTGATCTTC	1060
Qy	965	GCAGAGCGACACTTTATTTGGCACTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1024
Db	1061	GCAGAGCGACACTTTATTTGGCACTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1120
Qy	1025	AGCGGACCTCCAGGGAGCGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT	1084
Db	1121	AGCGGACCTCCATGGGAGCAGTCTGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT	1180
Qy	1085	GGGAGAGTCTTCTGATTTCTGGCCGGAGCACCAGACCTCTGATCTCTGGAGGAC	1144
Db	1181	GGGATGAATTTTGTATCTTCTGATCTTCTGATCTTCTGATCTTCTGATCTTCTG	1218
Qy	1145	TGAGTGTGCTCTTCTGAGGAGGCTGAGGAGGCTGAGGAGGCTGAGGAGGCTGAG	1203
Db	1219	-----ACCTCAAGCCAGGCTTGGCTGCTGAGGATGCTCTTCTTCTTCTTCTTCT	1270
Qy	1204	CTTCTCACTCCCAAGATGAGAGAGTGGCAGGAGCCCTCAAGGAGGAGAGCGCTCC	1263
Db	1271	TTTCTTTCTGCGAGAGGTGAGAGCAGTACCAAGAGGCTCTCTGGAAGAGGCAATTC	1330
Qy	1264	AT 1265	
Db	1331	AT 1332	
RESULT 4			
AK044629			
LOCUS			
DEFINITION			
Mus musculus adult retina cDNA, RIKEN full-length enriched library, clone:A930028F21 product:protein O-fucosyltransferase 1, full insert sequence.			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
1			
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,			
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.			
Normalization and subtraction of cap-trapper-selected cDNAs to			
prepare full-length cDNA libraries for rapid discovery of new genes			
Genome Res. 10 (10), 1617-1630 (2000)			
2			
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,			
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.			
Normalization and subtraction of cap-trapper-selected cDNAs to			
prepare full-length cDNA libraries for rapid discovery of new genes			
Genome Res. 10 (10), 1617-1630 (2000)			
3			
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,			

Konno, H., Akiyama, J., Nishi, K., Kitsunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

11076861

4

The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.

Functional annotation of a full-length mouse cDNA collection

Nature 409, 685-690 (2001)

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.

Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

6 (bases 1 to 2616)

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

## Direct Submission

## TITLE

## JOURNAL

Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gs.c.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

## COMMENT

cDNA library was prepared and sequenced in Mouse Genome

Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Retina RNA was provided by Dr. Stefano Gustincich (Department of Neurobiology, Harvard Medical School, 220 Longwood Ave., Boston, MA 02115, USA) whose assistance is gratefully acknowledged. Please visit our web site for further details.

URL: http://genome.gsc.riken.jp/

URL: http://fantom.gsc.riken.jp/

## FEATURES

## source

Location/Qualifiers  
1..2616  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="FANTOM DB:A930028F21"  
/db\_xref="taxon:10090"  
/clone="A930028F21"  
/tissue\_type="retina"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="adult"  
34..1215

## CDS

/note="unnamed protein product; protein  
O-fucosyltransferase 1 (MGD|GI:2153207, GB|NM\_080463,  
evidence: BLASTN, 99%, match=1290)  
putative"  
/codon\_start=1  
/protein\_id="BAC32009.1"  
/db\_xref="GI:26336653"

/translation="WGAAWAAPPHLLLRASFLLLLLPLRGRSAGSMDLACGLVLYCP  
CMGRGNOADHFLASLAFKLLNRLTAVPMLEYOHKPPFTNLHVSQYKPKLPLQ  
AYHRVVSLEPMENLAPSHMPPEKRYACFEVAQRSDDKTCPKMKNPFPFPDQF  
HVSFNKSLFTIGTISFSASYKEQWTFKPAQFPAQFVLEHRELRQITMV  
WSDENVRTEALISHLVVRPYVHLRISDWKNACAMKDGKTAGSHMFMSPOCVGYS  
RSTATPLMTMCLPDLKEIORAVTLWRALNARSVYIATDSSEYVSEIQOLFQKVRV  
VSLKPEVAQIDLYILGQADHFTGNCVSSFTAFVKRERDLHGRQSSFFGMDRPSQLRDE  
F"

## ORIGIN

Query Match 58.7%; Score 888; DB 3; Length 2616;

Best Local Similarity 87.1%; Pred. No. 7.4e-235;

Matches 975; Conservative 0; Mismatches 145; Indels 0; Gaps 0;

QY 3 GCCCGCGGCTCTGGGACCCGCGGTACTGCTCTACTGCCCTGCATGGCGGCTT 62  
DB 120 GTACAGCGGCTCTGGGACCTGGCGGTACTGCTCTACTGCTCTGATGGCGGCTT 179  
QY 63 TGGGAACAGGCGGATCACTTCTTGGGCTCTTGGCATTTGCAAGCTGCTAAACCGTAC 122  
DB 180 TGGGAACAGGCGGATCACTTCTTGGGCTCTTGGCATTTGCAAGCTGCTAAACCGTAC 239  
QY 123 CTGGCTGCTCTCTTGGATGAGTACAGCATCACAAAGCTCTCTTCCCAACCTCCA 182  
DB 240 CTGGCTGATCTCTTGGATGAGTACAGCATCACAAAGCTCTCTTCCCAACCTCCA 299  
QY 183 TGTGCTCTACAGAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCAGTCCGCTCATCAG 242  
DB 300 TGTGCTCTACCAAGTACTTCAAACTGGAGCTCTTCCAAAGCTTACCAGTCCGCTCATCAG 359  
QY 243 CTGGAGGATTTCTATGGAGAGAGTGGCAACCCACCTGGCCCTCGAGAACGGGTGGC 302  
DB 360 CTGGAGGATTTCTATGGAGAGAGTGGCAACCCACCTGGCCCTCGAGAACGGGTGGC 419  
QY 303 ATATGCTTTGAGGTGGCAGCCAGCAAGCCAGATAGAGAGCTGCCCATGAAGA 362  
DB 420 ATATGCTTTGAGGTGGCAGCCAGCAAGCCAGATAGAGAGCTGCCCATGAAGA 479  
QY 363 AGGAACCCCTTTGGCCCATTTCTGGATCAGTTTTCATGATGATTTCAACAGTCCGAGCT 422  
DB 480 AGGAACCCCTTTGGCCCATTTCTGGATCAGTTTTCATGATGATTTCAACAGTCCGAGCT 539  
QY 423 TTTTACAGGCTTTCTTCTCAGTGTCTTCTACAGAGAACCAATGGAGCCAGAGATTTTCTCC 482  
DB 540 GTTACAGGCTTTCTTCTCAGGCTCTCTACAGAGAACCAATGGAGCCAGAGATTTCTCC 599  
QY 483 AAAGAACATCGGTGCTTGGCTGCCAGAGCCAGCCAGTTCCTCGCTCTAGAGA 542  
DB 600 AAAAGAGCATCTGTGCTCGCACTGCTGGGCGCCCGCAGCAGTTCCTGTCTCGTGGAGA 659  
QY 543 ACAGAGCCACTACAGAGTACATGATGCTCAGACCAATGGTGAAGACGGGAGAGGC 602  
DB 660 ACAGAGGAGCTCCAGAAAGTACATGATGCTGATGATGATGATGATGATGATGATGATGAT 719  
QY 603 CCAGATTCATGCCCACTTGTTCGCGCCCTATGTGGGCTTTCATGCGCATTTGGCTCTGA 662  
DB 720 CTTGATCAGTTCCCACTCTGTGCGGCTCTATGTGGGCTTTCATGCGCATTTGGCTCTGA 779  
QY 663 CTGAAGAACCGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 722  
DB 780 CTGAAGAACCGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 839  
QY 723 TCCGAGTGTGCGGCTACAGCCGAGCAGCGGCGCCCTCTCAGATGATGATGATGATGATGATGAT 782  
DB 840 CCCTCAGTGTGCGGCTATAGCCGAGCAGCGGCGCCCTCTCAGATGATGATGATGATGATGATGAT 899  
QY 783 GCCTGACCTGAAGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGGTGAGGTGAGGTGAGGTGAGGT 842  
DB 900 CCCTGACCTGAAGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGGTGAGGTGAGGTGAGGTGAGGT 959  
QY 843 GTCGGTCTACTGCTTACTGATTCGAGAGTTATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 902  
DB 960 ATCGGTCTTACATCGCCACAGACTCTGAGAGCTACGTGTGAGAGATCCAGAGCTCTTCAA 1019



Tissue Procurement: DCTD/DTP  
CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone Distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNLNL at:  
<http://image.lnl.gov>

QY 1356 CAAAGCAGTCCAGCCTCGTCTTCTGGTCCACTCTGCTCTGACAGCCTGGGATGCTCAA 1415

Db 721 CAAGCAGTCAGCCTCCGTCCTTCTGGTCCACTCTGCTCTGAGCAGCCTGGGATGCTGAA 780

Qy	1416	CTCTTCAGAGAGATTTTTTTTATAGAGAGATTCTATATAATTTTGATACAAGGTCATGACTA	14175
Db	781	CTCTCTTCAGAGAGATTTTTTTTATAGAGAGA-TTCTATAAATTTTGATACAAGGTCATGACTA	839
Qy	1476	TCCTAGAACTCTCTGTGGTTTTTTTGAAA	1502
Db	840	TCCTAGAACTCTCTGTGNTTTTTTGAAA	866

ACCESSION	BM909055
VERSION	PM000055 1 CT-10350424

KEYWORDS  
EST.  
SOURCE  
ORGANISM  
Homo sapiens (human)  
Homo sapiens  
Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:

REFERENCE  
1 (bases 1 to 1091)  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

AUTHORS	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)

UNPUBLISHED (1999)  
Contact: Robert Strausberg, Ph.D.  
COMMENT

Email: [cs94008-remail@um.hk](mailto:cs94008-remail@um.hk)  
**Tissue Procurement: ATCC**

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

Clone distribution: MGC clone distribution information can be

http://image.llnl.gov

High quality sequence stop: 629.

```
source
1. .1091
/organism="Homo sapiens"
```

```
/mol_type="mrna"  
/db_xref="taxon:9606"
```

```
/clone="IMAGE:5431669"  
/tissue type="astrocytoma grade IV. cell line"
```

```
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 98"
```

/note="Organ: brain; Vector: pOTB7; Site1: XhoI; Site2: EcoRI; cDNA made by oligo-dT priming. Directionally

cloned into ECORI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong

in the laboratory of Gerard M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit

Note: this is a NIH\_MGC Library."

1  
 2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525

Best Local Similarity 93.0%; Pred. No. 7.4e-195;

Year	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035	2036	2037	2038	2039	2040	2041	2042	2043	2044	2045	2046	2047	2048	2049	2050	2051	2052	2053	2054	2055	2056	2057	2058	2059	2060	2061	2062	2063	2064	2065	2066	2067	2068	2069	2070	2071	2072	2073	2074	2075	2076	2077	2078	2079	2080	2081	2082	2083	2084	2085	2086	2087	2088	2089	2090	2091	2092	2093	2094	2095	2096	2097	2098	2099	2100																																																																																																																																																
1970	1.70	1.71	1.72	1.73	1.74	1.75	1.76	1.77	1.78	1.79	1.80	1.81	1.82	1.83	1.84	1.85	1.86	1.87	1.88	1.89	1.90	1.91	1.92	1.93	1.94	1.95	1.96	1.97	1.98	1.99	2.00	2.01	2.02	2.03	2.04	2.05	2.06	2.07	2.08	2.09	2.10	2.11	2.12	2.13	2.14	2.15	2.16	2.17	2.18	2.19	2.20	2.21	2.22	2.23	2.24	2.25	2.26	2.27	2.28	2.29	2.30	2.31	2.32	2.33	2.34	2.35	2.36	2.37	2.38	2.39	2.40	2.41	2.42	2.43	2.44	2.45	2.46	2.47	2.48	2.49	2.50	2.51	2.52	2.53	2.54	2.55	2.56	2.57	2.58	2.59	2.60	2.61	2.62	2.63	2.64	2.65	2.66	2.67	2.68	2.69	2.70	2.71	2.72	2.73	2.74	2.75	2.76	2.77	2.78	2.79	2.80	2.81	2.82	2.83	2.84	2.85	2.86	2.87	2.88	2.89	2.90	2.91	2.92	2.93	2.94	2.95	2.96	2.97	2.98	2.99	3.00	3.01	3.02	3.03	3.04	3.05	3.06	3.07	3.08	3.09	3.10	3.11	3.12	3.13	3.14	3.15	3.16	3.17	3.18	3.19	3.20	3.21	3.22	3.23	3.24	3.25	3.26	3.27	3.28	3.29	3.30	3.31	3.32	3.33	3.34	3.35	3.36	3.37	3.38	3.39	3.40	3.41	3.42	3.43	3.44	3.45	3.46	3.47	3.48	3.49	3.50	3.51	3.52	3.53	3.54	3.55	3.56	3.57	3.58	3.59	3.60	3.61	3.62	3.63	3.64	3.65	3.66	3.67	3.68	3.69	3.70	3.71	3.72	3.73	3.74	3.75	3.76	3.77	3.78	3.79	3.80	3.81	3.82	3.83	3.84	3.85	3.86	3.87	3.88	3.89	3.90	3.91	3.92	3.93	3.94	3.95	3.96	3.97	3.98	3.99	4.00	4.01	4.02	4.03	4.04	4.05	4.06	4.07	4.08	4.09	4.10	4.11	4.12	4.13	4.14	4.15	4.16	4.17	4.18	4.19	4.20	4.21	4.22	4.23	4.24	4.25	4.26	4.27	4.28	4.29	4.30	4.31	4.32	4.33	4.34	4.35	4.36	4.37	4.38	4.39	4.40	4.41	4.42	4.43	4

[illegible]



Qy 238 ATCAGCTTGGAGGATTTTCATGGAGAGCTGGCAACCACCACTGGCCCCCTCGAAGCGG 297  
Db |||||  
Qy 192 ATCAGCTTGGAGGATTTTCATGGAGAGCTGGCAACCACCACTGGCCCCCTCGAAGCGG 251  
Db |||||  
Qy 298 GTGGCATATCTCTTTGAGGTGGCAGCCAGCAAGCCAGATAGAGAGACGTGCCCCCATG 357  
Db |||||  
Qy 252 GTGGCATATCTCTTTGAGGTGGCAGCCAGCAAGCCAGATAGAGAGACGTGCCCCCATG 311  
Db |||||  
Qy 358 AAGGAAGAAACCCCTTTGGCCCATTTCTGGGATCAGTTTTCATGTGAGTTTCAACAAGTCG 417  
Db |||||  
Qy 312 AAGGAAGAAACCCCTTTGGCCCATTTCTGGGATCAGTTTTCATGTGAGTTTCAACAAGTCG 371  
Db |||||  
Qy 418 GAGCTTTTACAGGATTTCTTCAGTCTCTTACAGAGAACTATGGAGCCAGAGATTT 477  
Db |||||  
Qy 372 GAGCTTTTACAGGATTTCTTCAGTCTCTTACAGAGAACTATGGAGCCAGAGATTT 431  
Db |||||  
Qy 478 TCTCAAAAGAAACATCCGCTGCTTGGCCCTGCGAGAGCCAGCCAGATTCCTCCCTCTTA 537  
Db |||||  
Qy 432 TCTCAAAAGAAACATCCGCTGCTTGGCCCTGCGAGAGCCAGCCAGATTCCTCCCTCTTA 491  
Db |||||  
Qy 538 GAAGAAACACAGCCCACTACAGAAAGTACATGTATGTGTGAGCAAAATGGTGAAGACGGGA 597  
Db |||||  
Qy 492 GAGGAACACAGCCCACTACAGAAAGTACATGTATGTGTGAGCAAAATGGTGAAGACGGGA 551  
Db |||||  
Qy 598 GAGGCCAGATTCATGCCCCACCTTGTGCGGCCCTATGTGGGCAATTCATCTGCGCATTCGC 657  
Db |||||  
Qy 552 GAGGCCAGATTCATGCCCCACCTTGTGCGGCCCTATGTGGGCAATTCATCTGCGCATTCGC 611  
Db |||||  
Qy 658 TCTGACTGGAAGACGCTCTGTCATGCTGAAGACGGGACTGAGGCTCGCACTTCATG 717  
Db |||||  
Qy 612 TCTGACTGGAAGACGCTCTGTCATGCTGAAGACGGGACTGAGGCTCGCACTTCATG 671  
Db |||||  
Qy 718 GCCTCTCGCGAGTGTGGGTACAGCCGACAGCAGCGCCCTCTCACGATGACTATG 777  
Db |||||  
Qy 672 GCCTCTCGCGAGTGTGGGTACAGCCGACAGCAGCGCCCTCTCACGATGACTATG 731  
Db |||||  
Qy 778 TGCCCTGCTGACCTGAAGAGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTGCGTGGAT 837  
Db |||||  
Qy 732 TGCCCTGCTGACCTGAAGAGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTGCGTGGAT 791  
Db |||||  
Qy 838 GCCAGTGGCTTACGTTGCTACTGATTCAGAGAGTATGCTGAGTTCAGCTCCACAGCTC 897  
Db |||||  
Qy 792 GCCAGTGGCTTACGTTGCTACTGATTCAGAGAGTATGCTGAGTTCAGCTCCACAGCTC 851  
Db |||||  
Qy 898 TTCAAA--GGGAAGGTGAAGGTGGTG--AGCCTGAAGCTGAGGTGGCCAGGTGCACT 953  
Db |||||  
Qy 852 TTAAAGGGAGGGGAGGGTGGTGAACCTGGAACCTGTAGGGGGCCAGGTCCACCT 911  
Db |||||  
Qy 954 GTACATCTCGGCCCAAGCCGACCACTTT--ATTGGCACTGTGTCTCTCTCTCACTG--- 1009  
Db |||||  
Qy 912 GTAAATTTCTCGCAAGCCGAAACCTTTAAATTGGGAACCTGTGGCCCTCTCTCCGGGC 971  
Db |||||  
Qy 1010 CTTTGTGAAGCGGAGCGGACCTCCAGGGGAGGCGCTCTTT 1055  
Db |||||  
Qy 972 CTTTGGAAAGGAACGGGACCTCTCAGGGAGGGGCGCTTT 1017  
Db |||||

RESULT 8  
BX345989 841 bp mRNA linear EST 23-APR-2004  
LOCUS  
DEFINITION  
BX345989 Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED  
Homo sapiens cDNA clone CS0DJ008YF12 5-PRIME, mRNA sequence.  
ACCESSION  
BX345989.2 GI:46549776  
VERSION  
EST.  
KEYWORDS  
Homo sapiens (human)  
SOURCE  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 841)  
Li W.B., Gruber C., Jesse J. and Polayes D.  
AUTHORS  
Full-length cDNA libraries and normalization  
TITLE

Unpublished (2001)  
On May 5, 2003 this sequence version replaced gi:30377055.  
Genoscope - Centre National de Sequencage  
Contact: Genoscope  
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE  
Email: segre@genoscope.cns.fr, Web : www.genoscope.cns.fr  
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
end enriched, double-strand cDNA was digested with Not I and cloned  
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
was normalized. Library was constructed by Life Technologies, a  
division of Invitrogen. This sequence belongs to sequence cluster  
9580.f  
For more information about this cluster, see  
http://www.genoscope.cns.fr/cdna?s=CS0AJ008DC06QPl&c=9980.f.  
Location/Qualifiers  
1..841  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DJ008YF12"  
/cell\_type="T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED"  
/cell\_line="JURKAT"  
/clone\_lib="Homo sapiens T CELLS (JURKAT CELL LINE) COT  
10-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo(dT)  
primer. Five prime end enriched, double-strand cDNA was  
digested with Not I and cloned into the Not I and EcoR V  
sites of the pCMVSPORT 6 vector. Library was normalized."  
ORIGIN  
Query Match 47.5%; Score 718.4; DB 5; Length 841;  
Best Local Similarity 99.5%; Pred. No. 6.9e-188;  
Matches 730; Conservative 1; Mismatches 2; Indels 1; Gaps 1;  
Qy 1 ATGCCCGCGGGCTCTGGGACCCGCGGTTACTGCTTACTGCCCCCTCATGGGGCGC 60  
Db |||||  
Qy 109 ATGCTGCGGGCTCTGGGACCCGCGGTTACTGCTTACTGCCCCCTCATGGGGCGC 168  
Db |||||  
Qy 61 TTTGGGAACAGCGCGATCACTTCTTGGGCTCTTGGCATTTGCAAGCTGCTAAACCGT 120  
Db |||||  
Qy 169 TTTGGGAACAGCGCGATCACTTCTTGGGCTCTTGGCATTTGCAAGCTGCTAAACCGT 228  
Db |||||  
Qy 121 ACCTTGGCTGCTCCCTTGGATGAGTACCAAGCTACAGAGCTCTTTCACCACTC 180  
Db |||||  
Qy 229 ACCTTGGCTGCTCCCTTGGATGAGTACCAAGCTACAGAGCTCTTTCACCACTC 288  
Db |||||  
Qy 181 CATGTGCTTACAGAGTACTTCAAGCTGGAGCCCTTCCAGGCTTACCATGGGTCTATC 240  
Db |||||  
Qy 289 CATGTGCTTACAGAGTACTTCAAGCTGGAGCCCTTCCAGGCTTACCATGGGTCTATC 348  
Db |||||  
Qy 241 AGCTTGGAGGATTTTCATGGAGAGCTGGCACCCACCTCCACTGGCCCCCTGAGAGCGGGTG 300  
Db |||||  
Qy 349 AGCTTGGAGGATTTTCATGGAGAGCTGGCACCCACCTCCACTGGCCCCCTGAGAGCGGGTG 408  
Db |||||  
Qy 301 GCATATGCTTTGAGTGGCAGCCAGCAAGCCAGATAAGAGAGCTGCCCATGAAG 360  
Db |||||  
Qy 409 GCATATGCTTTGAGTGGCAGCCAGCAAGCCAGATAAGAGAGCTGCCCATGAAG 467  
Db |||||  
Qy 361 GAAGGAACCCCTTTGGCCCATTTCTGGGATCAGTTTCATGAGTTTCAACAGTGGAG 420  
Db |||||  
Qy 468 GAAGGAACCCCTTTGGCCCATTTCTGGGATCAGTTTCATGAGTTTCAACAGTGGAG 527  
Db |||||  
Qy 421 CTTTTCACAGGCAATTCCTTCAGTGTCTTCTACAGAGAACATGGAGCCAGAGATTTTCT 480  
Db |||||  
Qy 528 CTTTTCACAGGCAATTCCTTCAGTGTCTTCTACAGAGAACATGGAGCCAGAGATTTTCT 587  
Db |||||  
Qy 481 CCAAGGAACATCCGCTTGGCCCTGCGAGAGCCAGCCAGCTTCCCGTCTTAGAA 540  
Db |||||  
Qy 588 CCAAGGAACATCCGCTTGGCCCTGCGAGAGCCAGCCAGCTTCCCGTCTTAGAG 647  
Db |||||  
Qy 541 GRACAGGCCACTACAGAGTACATGTATGTGTGAGCAAGAAATGGTGAAGACGGGAG 600  
Db |||||  
Qy 648 GAACAGGCCACTACAGAGTACATGTATGTGTGAGCAAGAAATGGTGAAGACGGGAG 707  
Db |||||

JOURNAL  
COMMENT

FEATURES  
source



```

QY 601 GCCCAGATTATGCCCACTTGTTCGGCCCTATGTGGCCATTCATCTCGCATTTGGCTCT 660
Db 708 GSCCAGATTATGCCCACTTGTTCGGCCCTATGTGGCCATTCATCTCGCATTTGGCTCT 767
QY 661 GACTGGAAGAACGCTGTGCCATGCTGAAGACGGGACTGACGGTGCACATTCATGGCC 720
Db 768 GACTGGAAGAACGCTGTGCCATGCTGAAGACGGGACTGACGGTGCACATTCATGGCC 827
QY 721 TCTCCGAGTGTGT 734
Db 828 TCTCCGAGTGTGT 841

RESULT 9
LOCUS ALI135434
DEFINITION DKFZp76200115_r1 762 (synonym: hmel2) Homo sapiens cDNA clone
ACCESSION ALI135434
VERSION ALI135434.1 GI:6603621
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 719)
AUTHORS Ansorge, W., Wirkner, U., Mewes, H.W., Weil, B. and Wiemann, S.
TITLE EST (Ansorge, W., Wirkner, U., Mewes, H.W., Weil, B. and Wiemann, S.)
JOURNAL Unpublished (1999)
COMMENT Contact: MIPS
MIPS Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
Heidelberg/Germany) within the cDNA sequencing consortium of the
German Genome Project.
No sl sequence available.
This clone (DKFZp76200115) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES
    source
        1. 719
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="DKFZp76200115"
            /tissue type="melanoma (Mewo cell line)"
            /dev stage="adult"
            /lab_host="DH10B"
            /clone_lib="762 (synonym: hmel2)"
            /note="Vector: pSport1; Site_1: NotI; Site_2: SalI"

ORIGIN
Query Match 47.4%; Score 717.4; DB 1; Length 719;
Best Local Similarity 99.9%; Pred. No. 1.2e-187;
Matches 718; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 360 GGAAGGAACCCCTTTGGCCCATCTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGA 419
Db 1 GGAAGGAACCCCTTTGGCCCATCTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGA 60

QY 420 GCTTTTACAGGCATTTCTTTCAGTGTCTTCTCAGAGAACCAATGGAGCCAGAGATTTTC 479
Db 61 GCTTTTACAGGCATTTCTTTCAGTGTCTTCTCAGAGAACCAATGGAGCCAGAGATTTTC 120

QY 480 TCCAAAGAACATCCGGTGTTCCTCCAGAGAGCCCGCCAGCCAGTTCCCGCTCCTAGA 539
Db 121 TCCAAAGAACATCCGGTGTTCCTCCAGAGAGCCCGCCAGCCAGTTCCCGCTCCTAGA 180

QY 540 AGAACACAGGCCCACTACAGAAAGTACATGGTATGCTCAGACGCAAAATGGTGAAGACGGGAGA 599

```

```

Db 181 GGAACACAGGCCCACTACAGAAAGTACATGGTATGCTCAGACGAAATGGTGAAGACGGGAGA 240
QY 600 GGGCCAGATTATCGCCACCTTGTTCGGCCCTATGTGGGCAATTCATCTCGCATTTGGCTC 659
Db 241 GGGCCAGATTATCGCCACCTTGTTCGGCCCTATGTGGGCAATTCATCTCGCATTTGGCTC 300
QY 660 TGACTGGAAGAACCCCTTGTGCCATGCTGAAGACGGGACTGACGGTGCACATTCATGGC 719
Db 301 TGACTGGAAGAACCCCTTGTGCCATGCTGAAGACGGGACTGACGGTGCACATTCATGGC 360
QY 720 CTCTCCGAGTGTGTGGCTTACAGCCGAGACAGCGGCCCTTCCAGTACATATGTG 779
Db 361 CTCTCCGAGTGTGTGGCTTACAGCCGAGACAGCGGCCCTTCCAGTACATATGTG 420
QY 780 CTTCCCTGACTGAAGAGAGATCCAGAGGGCTGTCAGAGCTCTGGGTGAGGTGCTGGATGC 839
Db 421 CTTCCCTGACTGAAGAGAGATCCAGAGGGCTGTCAGAGCTCTGGGTGAGGTGCTGGATGC 480
QY 840 CCAGTCGGTCTACCTTGTCTACTGATTCGAGAGTTATGTGCTGAGCTCCCAACAGCTCTT 899
Db 481 CCAGTCGGTCTACCTTGTCTACTGATTCGAGAGTTATGTGCTGAGCTCCCAACAGCTCTT 540
QY 900 CAAAGGAAGTGAAGTGTGAGCTTGAAGCTTGAAGCTGAGGTGAGGTGAGTGCATGTACAT 959
Db 541 CAAAGGAAGTGAAGTGTGAGCTTGAAGCTTGAAGCTGAGGTGAGGTGAGTGCATGTACAT 600
QY 960 CTTCCGCAAGCCGACCACTTATTGCACTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1019
Db 601 CTTCCGCAAGCCGACCACTTATTGCACTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 660
QY 1020 GCGGAGCGGACCTCCAGGGAGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1078
Db 661 GCGGAGCGGACCTCCAGGGAGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 719

RESULT 10
LOCUS BI524031
DEFINITION 603052202P1 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5201616 5',
mRNA sequence.
ACCESSION BI524031
VERSION BI524031.1 GI:15348823
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 889)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11505 row: d column: 01
High quality sequence stop: 880.
FEATURES
    source
        1. 889
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="IMAGE:5201616"
            /lab_host="DH10B"
            /clone_lib="NIH_MGC_122"
            /note="Organ: pooled lung and spleen; Vector: pCMV-SPORT6;
            Site_1: NotI; Site_2: EcoRV (destroyed); RNA source

```

anonymous pool of 24 week female lung, 16 week female spleen, and 20-22 week male spleens. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.4 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 026. Note: this is a NIH\_MGC Library."

ORIGIN

Query Match 47.0%; Score 712.2; DB 4; Length 889;  
Best Local Similarity 99.1%; Pred. No. 3; 7e-106;  
Matches 758; Conservative 0; Mismatches 3; Indels 4; Gaps 4;

Qy 1 ATGCCCGGGGCTCTGGGACCGCGCGGTACTGCTCTACTGCGCCCTGCGATGGGGCG 60  
Db 87 ATGCCCGGGGCTCTGGGACCGCGCGGTACTGCTCTACTGCGCCCTGCGATGGGGCG 146  
Qy 61 TTGGGAACACGAGCGCATCACTTCTTGGGCTCTCTGGCATTTGCCAAGCTGTAACCGT 120  
Db 147 TTGGGAACACGAGCGCATCACTTCTTGGGCTCTCTGGCATTTGCCAAGCTGTAACCGT 206  
Qy 121 ACCTTGGCTGTCCCTCTCTGGATTGAGTACGAGCATCAAGGCTCTCTTTACCAAACTC 180  
Db 207 ACCTTGGCTGTCCCTCTCTGGATTGAGTACGAGCATCAAGGCTCTCTTTACCAAACTC 266  
Qy 181 CATGTGCTCTACCAAGTACTTCAAGCTGGAGCCCTCTCAGGCTTTACCATCGGTGATC 240  
Db 267 CATGTGCTCTACCAAGTACTTCAAGCTGGAGCCCTCTCAGGCTTTACCATCGGTGATC 326  
Qy 241 AGCTTGGAGGATTTTCATGGAGAGCTGGCACCACCCACTGGCCCTCGAGAAGCGGGT 300  
Db 327 AGCTTGGAGGATTTTCATGGAGAGCTGGCACCACCCACTGGCCCTCGAGAAGCGGGT 386  
Qy 301 GCATCTCTTTGAGTGGGAGCCGAGCCAGCCAGATGAAGAGCTGCCCATGAG 360  
Db 387 GCATCTCTTTGAGTGGGAGCCGAGCCAGCCAGATGAAGAGCTGCCCATGAG 446  
Qy 361 GAAGAAACCCCTTTGGCCCATTTCTGGGATCAGTTTCATGTGAGTTTCAACAGTCGGAG 420  
Db 447 GAAGAAACCCCTTTGGCCCATTTCTGGGATCAGTTTCATGTGAGTTTCAACAGTCGGAG 506  
Qy 421 CTTTTACAGCAATTTCTTCACTGCTCTTACAGAAACAATGGAGCCAGAGATTTTCT 480  
Db 507 CTTTTACAGCAATTTCTTCACTGCTCTTACAGAAACAATGGAGCCAGAGATTTTCT 566  
Qy 481 CCAAGGAACATCCGGTGTCTTGCCTTGGCCAGGAGCCCGAGTTCCTCCCTCTAGAA 540  
Db 567 CCAAGGAACATCCGGTGTCTTGCCTTGGCCAGGAGCCCGAGTTCCTCCCTCTAGAG 626  
Qy 541 GAACACGGCCACTACAGAGTACATGTATGTCAGACGAATGTGAAGACGGAGAG 600  
Db 627 GAACACGGCCACTACAGAGTACATGTATGTCAGACGAATGTGAAGACGGAGAG 686  
Qy 601 GCCAGATTGATGCCACCTGTGCCGCGC-CTATGTGGCAATTCATCTGCCATTTGCTC 659  
Db 687 GCCAGATTGATGCCACCTGTGCCGCGC-CTATGTGGCAATTCATCTGCCATTTGCTC 746  
Qy 660 TGACT-GGAGAAGCCCTGTGCCATGTCTGAGGAGCGG-ACTCGAGCTTCGACTTCATG 717  
Db 747 TGACTGGGAAGAGCCCTGTGCCATGTCTGAGGAGCGGAACTCGAGCTTCGACTTCATG 806  
Qy 718 GCCTCTCCGCAAGTGTGGGCTACAGCCGACGACAGCGGCCCC 762  
Db 807 GCCTCTCCGCAAGTGTGGGCTACAG-CCGACGACAGCGGCCCC 850

RESULT 11  
BX449172  
LOCUS  
DEFINITION BX449172 Homo sapiens FETAL LIVER Homo sapiens cDNA clone  
CS0DM014YK11 5-PRIME, mRNA sequence.  
ACCESSION BX449172

VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

BX449172.2 GI:47067286  
EST.  
Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 898)  
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
Unpublished (2001)  
On May 22, 2003 this sequence version replaced gi:31030014.  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE  
Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr  
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
end enriched, double-strand cDNA was digested with Not I and cloned  
into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library  
was not normalized. Library was constructed by Life Technologies, a  
division of Invitrogen.  
This sequence belongs to sequence cluster 9980.f  
For more information about this cluster, see  
http://www.genoscope.cns.fr/cdna?s=CS0AM014AF06QPl&c=9980.f.  
Location/Qualifiers  
1..898  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DM014YK11"  
/tissue\_type="FETAL LIVER"  
/dev\_stage="fetal"  
/clone\_lib="Homo sapiens FETAL LIVER"  
/notes="Organ: liver; Vector: pCMVSPORT 6; 1st strand cDNA  
was primed with a NotI-oligo(dT) primer; Five prime end  
enriched, double-strand cDNA was digested with Not I and  
cloned into the Not I and EcoRV sites of the pCMVSPORT 6  
vector. Library was not normalized."

FEATURES  
source

Query Match 46.7%; Score 706.6; DB 5; Length 898;  
Best Local Similarity 97.8%; Pred. No. 1.3e-184;  
Matches 744; Conservative 4; Mismatches 10; Indels 3; Gaps 3;

Qy 1 ATGCCCGGGGCTCTGGGACCGCGCGGTACTGCTCTACTGCGCCCTGCGATGGGGCG 60  
Db 139 ATGCCCGGGGCTCTGGGACCGCGCGGTACTGCTCTACTGCGCCCTGCGATGGGGCG 198  
Qy 61 TTTGGGAACACGAGCGCATCACTTCTTGGGCTCTCTGGCATTTGCCAAGCTGCTAAACCGT 120  
Db 199 TTTGGGAACACGAGCGCATCACTTCTTGGGCTCTCTGGCATTTGCCAAGCTGCTAAACCGT 258  
Qy 121 ACCTTGGCTGTCCCTCTCTGGATTGAGTACGAGCATCAAGGCTCTCTTTCACCAACCTC 180  
Db 259 ACCTTGGCTGTCCCTCTCTGGATTGAGTACGAGCATCAAGGCTCTCTTTCACCAACCTC 318  
Qy 181 CATGTGCTCTACGAGTACTTTCAGCTGGAGCCCTCCAGGCTTACCATCGGTGATC 240  
Db 319 CATGTGCTCTACGAGTACTTTCAGCTGGAGCCCTCCAGGCTTACCATCGGTGATC 378  
Qy 241 AGCTTGGAGGATTTTCATGGAGAGCTGGCACCACCCACTGGGCCCCCTGAGAAGCGGGT 300  
Db 379 AGCTTGGAGGATTTTCATGGAGAGCTGGCACCACCCACTGGGCCCCCTGAGAAGCGGGT 438  
Qy 301 GCATCTCTTTGAGGTGGCAGCCCGAGGAGCCAGATGAAGAGCTGCCCCCATGAGAG 360  
Db 439 GCATCTCTTTGAGGTGGCAGCCCGAGGAGCCAGATGAAGAGCTGCCCCCATGAGAG 497  
Qy 421 CTTTTACAGCAATTTCTTCACTGCTCTTACAGAGAAACAATGGAGCCAGAGATTTTCT 480

ORIGIN

```

Db      558 CTTTTCACAGGCAATTCCTTCAGTGTTCGTACAGAGAA CAATGGAGCCAGAGATTTTCT 617
Qy      481 CCAAGGAAACATCCGGTGTCTTCCCTGCCAGGAGCCCGCCAGCCAGTTCCTCCGTCCTAGAA 540
Db      618 CCNAGGAAACATCCGGTGTCTTCCCTGCCAGGAGCCCGCCAGCCAGTTCCTCCGTCCTAGAG 677
Qy      541 GAACACAGGCCACTACAGAAAGTACATGGTATGTGTACAGCG- AAATGGTGAAGACGGGAGA 599
Db      678 GAACACAGGCCACTACAGAAAGTACATGGTATGTGTACAGCGAAATGGTGAAGACGGGAGA 737
Qy      600 GGGCCAGATCATGCCCACTTGTCCGGCCCTATGTGGGCATTCATCTCGGATTTGGCTC 659
Db      738 CGCCCAAGTTCATGCCCCCTTGTCCGGCCCTATGTGGGCATTCATCTCGGATTTGGCTC 797
Qy      660 TGACTGGAAGAACCGCTGTGCCATGTGGAAGACGGGACTGCAGGCTCGCACTTCATGSC 719
Db      798 TGACTGGAAGAACCGCTGTCCATGTCTGAAGACGGGACTGCAGGCTCGCACTTCATGSC 857
Qy      720 CT-CTCCGAGTGTGGGCTACAGCCGACGACGCGCC 759
Db      858 CTYCTCCGAGTGTGGGCDACAGCCGACGACACGCGCC 898

RESULT 12
BM475212
LOCUS      704 bp mRNA linear EST 05-FEB-2002
DEFINITION AGENCOURT 6477457 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:5559347
5', mRNA sequence.
ACCESSION BM475212
VERSION   1
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 704)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
          Email: cgaabs-r@mail.nih.gov
          Tissue Procurement: ATCC
          cDNA Library Preparation: Life Technologies, Inc.
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
          DNA Sequencing by: Agencourt Bioscience Corporation
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: LLAM12283 row: m column: 12
          High quality sequence stop: 685.
          Location/Qualifiers
            1..704
              /organism="Homo sapiens"
              /mol_type="mRNA"
              /db_xref="taxon:9606"
              /clone="IMAGE:5559347"
              /tissue_type="duodenal adenocarcinoma, cell line"
              /lab_host="DH10B (phage-resistant)"
              /clone_lib="NIH_MGC_88"
              /note="Organ: small intestine; Vector: pCMV-SPORT6;
              Site 1: NotI; Site 2: SalI; Cloned unidirectionally;
              oligo-dT primed. Average insert size 1.767 kb. Library
              enriched for full-length clones and constructed by Life
              Technologies. Note: this is a NIH_MGC Library."

FEATURES
             source
             1..704
               /organism="Homo sapiens"
               /mol_type="mRNA"
               /db_xref="taxon:9606"
               /clone="IMAGE:5559347"
               /tissue_type="duodenal adenocarcinoma, cell line"
               /lab_host="DH10B (phage-resistant)"
               /clone_lib="NIH_MGC_88"
               /note="Organ: small intestine; Vector: pCMV-SPORT6;
               Site 1: NotI; Site 2: SalI; Cloned unidirectionally;
               oligo-dT primed. Average insert size 1.767 kb. Library
               enriched for full-length clones and constructed by Life
               Technologies. Note: this is a NIH_MGC Library."

ORIGIN
Query Match 45.2%; Score 683.8; DB 4; Length 704;
Best Local Similarity 99.7%; Pred. No. 2.7e-178;
Matches 685; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      797 AGATCCAGAGGCTGTGAAGCTCTCGGTGAGTCTCGTGTAGTCCGTCGATGCCAGTCCGTCACGTTG 856
Db      3 AGATCCAGAGGCTGTGAAGCTCTCGGTGAGTCTCGTGTAGTCCGTCGATGCCAGTCCGTCACGTTG 62

```

```

Qy      857 CTACTGATTCGAGAGTTATGTGCTGAGCTCAACAGAGCTCTTCAAAGGGAAGGTGAAG 916
Db      63 CTACTGATTCGAGAGTTATGTGCTGAGCTCAACAGAGCTCTTCAAAGGGAAGGTGAAG 122
Qy      917 TGGTGAAGCTGAAGCCTGAGGTGGCCAGGTGCAGCTGTATCATCTCTCGCCCAAGCGGACC 976
Db      123 TGGTGAAGCTGAAGCCTGAGGTGGCCAGGTGCAGCTGTATCATCTCTCGCCCAAGCGGACC 182
Qy      977 ACTTTATTTGGCAACTGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1036
Db      183 ACTTTATTTGGCAACTGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 242
Qy      1037 AGGGGAGGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1096
Db      243 AGGGGAGGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 302
Qy      1097 GATTTCTGGCCGAGCACCAGACCCCTCTGATCTCTGGAGGACCAGAGTCTGAGCTGTGCTCT 1156
Db      303 GATTTCTGGCCGAGCACCAGACCCCTCTGATCTCTGGAGGACCAGAGTCTGAGCTGTGCTCT 362
Qy      1157 TCCAGCCAGGCTGCGCAGCAGAGTGTCTCGGGATTGCAAACTCTCTCTCTCTCTCTCTCTCTCT 1216
Db      363 TCCAGCCAGGCTGCGCAGCAGAGTGTCTCGGGATTGCAAACTCTCTCTCTCTCTCTCTCTCTCT 422
Qy      1217 AAAGATGAGAAAGAGTGTCCAGGAGCCCTCAAGGAGGAGAGCGCTCCATATCCAGGGCA 1276
Db      423 AAAGATGAGAAAGAGTGTCCAGGAGCCCTCAAGGAGGAGAGCGCTCCATATCCAGGGCA 482
Qy      1277 TAGGACTTGCAGGTTCTTAGGAGCAGGAGCATCTCCCATCGCACGTGCTTCTCTCTCTCTCT 1336
Db      483 TAGGACTTGCAGGTTCTTAGGAGCAGGAGCATCTCCCATCGCACGTGCTTCTCTCTCTCTCT 542
Qy      1337 TGGGAATTTCTCAGACTGCGCAAGCAGTCCAGCCTCGCTCTCTCTCTCTCTCTCTCTCTCTCT 1396
Db      543 TGGGAATTTCTCAGACTGCGCAAGCAGTCCAGCCTCGCTCTCTCTCTCTCTCTCTCTCTCTCT 602
Qy      1397 AGCAGCTGGGATGCTGAACTCTTCAGAGAGATTTTTTTATAGAGAGATTTCTATAATTT 1456
Db      603 AGCAGCTGGGATGCTGAACTCTTCAGAGAGATTTTTTTATAGAGAGATTTCTATAATTT 662
Qy      1457 TGATACAAAGGTCATGACTATCTAGAA 1483
Db      663 TGATACAAAGGTCATGACTATCTAGAA 689

RESULT 13
CO774833 942 bp mRNA linear EST 04-AUG-2004
LOCUS      ILLUMIGEN MCQ 51950 Katze MNLV Macaca nemestrina cDNA clone
DEFINITION ILLUM:29950 5', similar to Bases 7 to 852 highly similar to human
          POFUT1 (Hs.178292), mRNA sequence.
ACCESSION CO774833
VERSION   1
KEYWORDS  EST.
SOURCE    Macaca nemestrina (pig-tailed macaque)
ORGANISM  Macaca nemestrina
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
Cercopithecinae; Macaca.
REFERENCE 1 (bases 1 to 942)
AUTHORS   Katze M.G., Thomas M., Korth M., Iadonato S.P. and Magness C.L.
TITLE     Large-scale Rhesus Macaque cDNA Sequencing
JOURNAL   Unpublished (2003)
COMMENT   Contact: C. Magness
          Illumigen Biosciences Inc.
          2203 Airport Way S, Suite 450, Seattle, WA 98134, USA
          Tel: 2063780400
          Fax: 2063780408
          Email: cmagness@illumigen.com
          Sequenced on 2004.07.21. 611 Q20 bases. Library Preparation: Prof.
          Michael Katze Lab at University of Washington DNA Sequencing:
          Illumigen Biosciences Inc. For further information, see

```





Search completed: November 9, 2005, 10:38:05  
Job time : 5495.13 secs

---

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2005, 23:53:44 ; Search time 4962.62 Seconds  
(without alignments)  
10740.434 Million cell updates/sec

Title: US-09-774-954-16  
 Perfect score: 1100  
 Sequence: 1 ATGCCCGGGGGCTCTGGGA.....CTCGGGGACGAGTTCGTGATT 1100

Scoring table: IDENTITY\_NUC  
Gapop 10.0 . Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:★

1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vi:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	1100	100.0	1100	6	AR163457	Sequence
2	1100	100.0	1100	6	BD103196	O-fucosyl
3	1100	100.0	1514	6	AR163451	Sequence
4	1100	100.0	1514	6	BD103188	O-fucosyl
5	1098.4	99.9	1300	6	AR163452	Sequence
6	1098.4	99.9	1300	6	BD103189	O-fucosyl
7	1098.4	99.9	5249	9	AF375884	Homo sapi
8	1098.4	99.9	11284	6	BD103190	O-fucosyl
9	1096.8	99.7	1189	9	D80002	Homo sapien
10	1096.4	99.7	1167	6	CQ727777	Sequence
11	1095.8	99.6	5266	6	AX780153	Sequence
12	1095.2	99.6	4560	9	AL832023	Homo sapi
13	1086.8	98.8	1167	9	AFJ781500	Pan trogl
14	1033.4	93.9	5009	6	AR163453	Sequence
15	1033.4	93.9	5009	6	BD103191	O-fucosyl
16	916.2	83.3	1176	4	AX344580	Bos tauru
17	908.2	82.6	1176	4	AFJ781503	Bos tauru
18	901.6	82.0	1134	4	SSC567917	Sus scrofa
19	889.6	80.9	1188	10	AFJ781499	Rattus norv

20	885.8	80.5	1302	10	AF375885	Mus muscu
21	885.8	80.5	1529	10	BC046295	Mus muscu
22	752.8	68.4	1056	4	AY344581	Bos tauru
23	652.2	59.3	1732	5	GA535754	Gallus ga
24	652.2	59.3	2546	5	AJ720352	Gallus ga
25	586.4	53.3	1502	5	BC082519	Xenopus t
26	586.4	53.3	1538	5	AJ781498	Silurana
27	563.6	51.2	1155	5	AJ606070	Fugu rubr
28	552.8	50.3	1505	5	XLA514425	Xenopus l
29	539	49.0	1164	5	AJ781504	Tetraodon
30	538.4	48.9	1549	5	BC579536	Danio rer
31	471.4	42.9	1524	5	BC000582	Homo sapi
32	445.4	40.5	920	5	AJ719585	Gallus ga
33	251.2	22.8	145414	9	HS392M18	Human DNA
34	201.6	18.3	1841	3	AK122708	Ciona int
35	193.4	17.6	2393	9	AK124093	Homo sapi
36	188.8	17.2	240336	2	AC094844	Rattus no
37	186.6	17.0	240510	2	AC134371	Rattus no
38	184.6	16.8	200329	10	AC078911	Mus muscu
39	184.6	16.8	208614	10	AL807380	Mouse DNA
40	184	16.7	1145	3	AJ781502	Ciona sav
41	172.8	15.7	1209	3	AJ831490	Drosophil
42	168	15.3	1295	3	AY118651	Drosophil
43	168	15.3	1320	6	AY809393	Sequence
44	166.8	15.2	1209	3	AJ781501	Drosophil
45	166.4	15.1	1209	6	C0590403	Sequence

## ALIGNMENTS

```

RESULT 1
AR163457
LOCUS
DEFINITION
AR163457
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
ORIGIN

```

Query Match	100.0%	Score 1100;	DB 6;	Length 1100;
Best Local Similarity	100.0%	Pred. No. 1e-253;		
Matches 1100;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	ATGCCCGCGGGTCTCTGGGACCGCGCGGTACTCTCTACTGCGCCCTGATGGGGCGC	60	
Db	1	ATGCCCGCGGGTCTCTGGGACCGCGCGGTACTCTCTACTGCGCCCTGATGGGGCGC	60	
Qy	61	TTTGGGAACAGGCGCGATCACTTCTTTGGGCTCTCTGGCATTTGCAAGCTGCTAAACCGT	120	
Db	61	TTTGGGAACAGGCGCGATCACTTCTTTGGGCTCTCTGGCATTTGCAAGCTGCTAAACCGT	120	
Qy	121	ACCTTGGGTGTCCTCTCTGGATTTAGTACACAGCATCACAGGCTCTTTTACCAACCTC	180	
Db	121	ACCTTGGGTGTCCTCTCTGGATTTAGTACACAGCATCACAGGCTCTCTTTTACCAACCTC	180	
Qy	181	CATGTGTCTTACAGAACTTCAAGCTGGAGCCCTCCAGGCTTTACCATCGGTCATC	240	
Db	181	CATGTGTCTTACAGAACTTCAAGCTGGAGCCCTCCAGGCTTTACCATCGGTCATC	240	
Qy	241	AGCTTGGAGGATTTTCATGGAGAAGCTGGACCCACCCACTGGCCCTCCCTGAGAAAGCGGGTG	300	

Db 241 ACCTTGGAGGATTTTCATGAGAACTGGACCCACCACCTGGCCCCCTGAGAACGGGTG 300  
Qy 301 GCATATCTCTTTGAGGTGGCGCCAGCGAAGCCAGATTAAGAACTGCCCCATGAAG 360  
Db 301 GCATATCTCTTTGAGGTGGCGCCAGCGAAGCCAGATTAAGAACTGCCCCATGAAG 360  
Qy 361 GAAGAAACCCCTTTGGCCCATCTTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420  
Db 361 GAAGAAACCCCTTTGGCCCATCTTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420  
Qy 421 CTTTTCACAGCATTTCTCTTCACTAGTCTTCTTACAGAAACATGGAGCCAGAGATTTTCT 480  
Db 421 CTTTTCACAGCATTTCTCTTCACTAGTCTTCTTACAGAAACATGGAGCCAGAGATTTTCT 480  
Qy 481 CCAAGGAACATCCGGTGTCTTGGCCCTGGCAGAGCCCGCCAGTTCGCCGCTCTAGAA 540  
Db 481 CCAAGGAACATCCGGTGTCTTGGCCCTGGCAGAGCCCGCCAGTTCGCCGCTCTAGAA 540  
Qy 541 GAACACAGCCACTACAGAACTACATGTTGTTGACAGCAAAATGTTGAAGACGGAGAG 600  
Db 541 GAACACAGCCACTACAGAACTACATGTTGTTGACAGCAAAATGTTGAAGACGGAGAG 600  
Qy 601 GCCAGATTCATGCCACCTTGTCCGGCCCTATGTGGGCAATTCATCTGGCATTTGGTCT 660  
Db 601 GCCAGATTCATGCCACCTTGTCCGGCCCTATGTGGGCAATTCATCTGGCATTTGGTCT 660  
Qy 661 GACTGGAAGAACCGCTGTGCATGCTGAAGACGGGACTGCGAGCTCGCACTTCATGSCC 720  
Db 661 GACTGGAAGAACCGCTGTGCATGCTGAAGACGGGACTGCGAGCTCGCACTTCATGSCC 720  
Qy 721 TCTCCGCAAGTGTGGGCTACAGCCGACACAGCGCCCGCCCTCACGATGACTATGTC 780  
Db 721 TCTCCGCAAGTGTGGGCTACAGCCGACACAGCGCCCGCCCTCACGATGACTATGTC 780  
Qy 781 CTGCTGACCTGAAGGAGATTCAGAGGGCTGTGAAGCTCTGGGTGAGTTCGGTATGCC 840  
Db 781 CTGCTGACCTGAAGGAGATTCAGAGGGCTGTGAAGCTCTGGGTGAGTTCGGTATGCC 840  
Qy 841 CAGTCCGTCTACGTGTCTACTGATTCAGAGAGTTATGTGCTGAGCTCCAAAGCTCTTC 900  
Db 841 CAGTCCGTCTACGTGTCTACTGATTCAGAGAGTTATGTGCTGAGCTCCAAAGCTCTTC 900  
Qy 901 AAAGGGAAGTGAAGTGTGGTGAAGCTGAGGTGGCCAGCTCCAGCTGACATC 960  
Db 901 AAAGGGAAGTGAAGTGTGGTGAAGCTGAGGTGGCCAGCTCCAGCTGACATC 960  
Qy 961 CTGCGCAAGCCGACCACTTTATGGCACTGTGTCTCTCTCACTGCTTTGTGAAG 1020  
Db 961 CTGCGCAAGCCGACCACTTTATGGCACTGTGTCTCTCTCACTGCTTTGTGAAG 1020  
Qy 1021 CGGAGCGGAGCTTCAGAGGGAGGCGCTCTCTTTCTTGGGATGAGACAGGCCCTTAAG 1080  
Db 1021 CGGAGCGGAGCTTCAGAGGGAGGCGCTCTCTTTCTTGGGATGAGACAGGCCCTTAAG 1080  
Qy 1081 CTGCGGAGCAGATTTCTGATT 1100  
Db 1081 CTGCGGAGCAGATTTCTGATT 1100

RESULT 2  
BD103196  
LOCUS BD103196 1100 bp DNA linear PAT 27-AUG-2002  
DEFINITION O-fucosyltransferase.  
ACCESSION BD103196  
VERSION BD103196.1 GI:22648770  
KEYWORDS JP 2001527389-A/9.  
SOURCE unidentified  
ORGANISM unidentified  
REFERENCE 1 (bases 1 to 1100)  
AUTHORS Wang, Y. and Spellman, M. W.  
TITLE O-fucosyltransferase  
JOURNAL Patent: JP 2001527389-A 9 25-DEC-2001;

COMMENT GENENTECH INC  
OS Unidentified  
PN JP 2001527389-A/9  
PD 25-DEC-2001  
PF 17-DEC-1997 JP 1998532877  
PR 31-JAN-1997 US 08/752498, 26-NOV-1997 US 08/978741 PI  
YANG WANG, MICHAEL W SPELLMAN  
PC C12N15/54, C12N9/10, C07K16/40  
CC Strandedness: Single;  
CC Topology: Linear;  
CC O-fucosyltransferase  
CC Location/Qualifiers  
FH Key 1..1100  
FT source /organism='Unidentified'.  
FEATURES  
source  
1..1100  
/organism='unidentified'  
/mol\_type='genomic DNA'  
/db\_xref='taxon:32644'  
ORIGIN  
Query Match 100.0%; Score 1100; DB 6; Length 1100;  
Best Local Similarity 100.0%; Pred. No. 1e-253;  
Matches 1100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 ATGCCCGCGGCTCTGGGACCGCGCGGTACTGCTCTACTGCGCTGCGATGGGGCG 60  
Db 1 ATGCCCGCGGCTCTGGGACCGCGCGGTACTGCTCTACTGCGCTGCGATGGGGCG 60  
Qy 61 TTTGGGAACACAGCGCGATCACTTTCTTGGGCTCTCTGGCATTTGCAAGCTGTAACCGT 120  
Db 61 TTTGGGAACACAGCGCGATCACTTTCTTGGGCTCTCTGGCATTTGCAAGCTGTAACCGT 120  
Qy 121 ACCTTGGGTGTCCTCTCTGGATGTAGTACACAGCATCAAGCTCTCTTTCACCAACCTC 180  
Db 121 ACCTTGGGTGTCCTCTCTGGATGTAGTACACAGCATCAAGCTCTCTTTCACCAACCTC 180  
Qy 181 CATGTGTCTTACCAAGAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGCTCATC 240  
Db 181 CATGTGTCTTACCAAGAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGCTCATC 240  
Qy 241 ACCTTGGAGGATTTTCATGAGAGTGGCAGCCACCCACCTGGCCCTCGAAGCGGGTG 300  
Db 241 ACCTTGGAGGATTTTCATGAGAGTGGCAGCCACCCACCTGGCCCTCGAAGCGGGTG 300  
Qy 301 GCATATCTCTTTGAGGTGGCGCCAGCGAAGCCAGATTAAGAACTGCCCCATGAAG 360  
Db 301 GCATATCTCTTTGAGGTGGCGCCAGCGAAGCCAGATTAAGAACTGCCCCATGAAG 360  
Qy 361 GAAGAAACCCCTTTGGCCCATCTTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420  
Db 361 GAAGAAACCCCTTTGGCCCATCTTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420  
Qy 421 CTTTTCACAGCATTTCTCTTCACTAGTCTTCTTACAGAAACATGGAGCCAGAGATTTTCT 480  
Db 421 CTTTTCACAGCATTTCTCTTCACTAGTCTTCTTACAGAAACATGGAGCCAGAGATTTTCT 480  
Qy 481 CCAAGGAACATCCGGTGTCTTGGCCCTGGCAGAGCCCGCCAGTTCGCCGCTCTAGAA 540  
Db 481 CCAAGGAACATCCGGTGTCTTGGCCCTGGCAGAGCCCGCCAGTTCGCCGCTCTAGAA 540  
Qy 541 GAACACAGCCACTACAGAACTACATGTTGTTGACAGCAAAATGTTGAAGACGGAGAG 600  
Db 541 GAACACAGCCACTACAGAACTACATGTTGTTGACAGCAAAATGTTGAAGACGGAGAG 600  
Qy 601 GCCAGATTCATGCCACCTTGTCCGGCCCTATGTGGGCAATTCATCTGGCATTTGGTCT 660  
Db 601 GCCAGATTCATGCCACCTTGTCCGGCCCTATGTGGGCAATTCATCTGGCATTTGGTCT 660  
Qy 661 GACTGGAAGAACCGCTGTGCATGCTGAAGACGGGACTGCGAGCTCGCACTTCATGSCC 720  
Db 661 GACTGGAAGAACCGCTGTGCATGCTGAAGACGGGACTGCGAGCTCGCACTTCATGSCC 720



QY 721 TCTCCGACGTGTGTGGCTTACAGCCGACAGCAGCGGCCCCCTCACAGATGACTATGTGC 780  
DB 721 TCTCCGACGTGTGTGGCTTACAGCCGACAGCAGCGGCCCCCTCACAGATGACTATGTGC 780  
QY 781 CTGCGCTGACCTGAAGAGATCAGAGGGCTGTGAAGCTCTGGTGAGGTTCGTGATGCC 840  
DB 781 CTGCGCTGACCTGAAGAGATCAGAGGGCTGTGAAGCTCTGGTGAGGTTCGTGATGCC 840  
QY 841 CAGTCGGTCTAGCTTCTACTGATTCGAGAGTTATGTGCTTCCAGCTTCCAAACAGCTCTTC 900  
DB 841 CAGTCGGTCTAGCTTCTACTGATTCGAGAGTTATGTGCTTCCAGCTTCCAAACAGCTCTTC 900  
QY 901 AAAGGAAGGTGAAGTGTGAGCTTGAAGCCTGAGGTGGCCAGCTGACCTGTACATC 960  
DB 901 AAAGGAAGGTGAAGTGTGAGCTTGAAGCCTGAGGTGGCCAGCTGACCTGTACATC 960  
QY 961 CTGCGCAGCGGACCACTTTATTTGGCAACTGTGTCTCCTCTTCACTGCTTGTGAAG 1020  
DB 961 CTGCGCAGCGGACCACTTTATTTGGCAACTGTGTCTCCTCTTCACTGCTTGTGAAG 1020  
QY 1021 CGGAGCGGACCTCCAGGGAGGGCGTCTTCTTTCTCGGATGAGCAGGCGCCCTAAG 1080  
DB 1021 CGGAGCGGACCTCCAGGGAGGGCGTCTTCTTTCTCGGATGAGCAGGCGCCCTAAG 1080  
QY 1081 CTGCGGACGAGTTCGTATT 1100  
DB 1081 CTGCGGACGAGTTCGTATT 1100

RESULT 3  
LOCUS AR163451 1514 bp DNA linear PAT 17-OCT-2001  
DEFINITION Sequence 2 from patent US 6270987.  
ACCESSION AR163451  
VERSION AR163451.1 GI:16234056  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknow.  
REFERENCE 1 (bases 1 to 1514)  
AUTHORS Wang, Y. and Spellman, M.W.  
TITLE O-fucosyltransferase  
JOURNAL Patent: US 6270987-A 2 07-AUG-2001;  
FEATURES  
source location/Qualifiers  
1..1514  
/organism="unknown"  
/mol\_type="unassigned DNA"

ORIGIN  
Query Match 100.0%; Score 1100; DB 6; Length 1514;  
Best Local Similarity 100.0%; Pred. No. 1e-253;  
Matches 1100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCCCCGGGGTCTCTGGGACCCGCGGGTTACCTGCTCTACTGCCCTGATGGGGCGC 60  
DB 1 ATGCCCCGGGGTCTCTGGGACCCGCGGGTTACCTGCTCTACTGCCCTGATGGGGCGC 60  
QY 61 TTTGGGAACCGGCGGATCTTCTGGGCTCTCTGGCACTTGTGCAATGCTGAACCGT 120  
DB 61 TTTGGGAACCGGCGGATCTTCTGGGCTCTCTGGCACTTGTGCAATGCTGAACCGT 120  
QY 121 ACCTTGGGTGTCCTCTTGGATTGAGTACAGCATCAAGACCTCTCTTTCACCAACCTC 180  
DB 121 ACCTTGGGTGTCCTCTTGGATTGAGTACAGCATCAAGACCTCTCTTTCACCAACCTC 180  
QY 181 CATGTGCTCTACAGAGTACTTCAAGCTGGAGGCCCTCCAGGCTTACATCGGCTCATC 240  
DB 181 CATGTGCTCTACAGAGTACTTCAAGCTGGAGGCCCTCCAGGCTTACATCGGCTCATC 240  
QY 241 AGCTTGGAGGATTTATGAGAGAGCTGGACCCACCACTGGCCCCCTGAGAGGGGTG 300  
DB 241 AGCTTGGAGGATTTATGAGAGAGCTGGACCCACCACTGGCCCCCTGAGAGGGGTG 300

QY 301 GCATATCTGCTTTGAGTGGCAGCCAGCGAAGCCAGATAAGAAAGACGTGCCCCCTGAAG 360  
DB 301 GCATATCTGCTTTGAGTGGCAGCCAGCGAAGCCAGATAAGAAAGACGTGCCCCCTGAAG 360  
QY 361 GAAGAAACCCCTTTGGCCCATTTCTGGGATCAATGTTTCAATGTTTCAACAGTCGGAG 420  
DB 361 GAAGAAACCCCTTTGGCCCATTTCTGGGATCAATGTTTCAATGTTTCAACAGTCGGAG 420  
QY 421 CTTTTTACAGGATTTCTTCAAGTCTTCTTACAGAACAAATGTTGAGCCAGAGATTTTCT 480  
DB 421 CTTTTTACAGGATTTCTTCAAGTCTTCTTACAGAACAAATGTTGAGCCAGAGATTTTCT 480  
QY 481 CCAAGGAACATCCCGTCTTCCCTGCGAGGAGCCAGCCAGCTTCCCTCTCTAGAA 540  
DB 481 CCAAGGAACATCCCGTCTTCCCTGCGAGGAGCCAGCCAGCTTCCCTCTCTAGAA 540  
QY 541 GAACACAGGCGCATACAGAAATGATGTTGTTGAGTGTGAGTGTGAGTGTGAGTGTGAG 600  
DB 541 GAACACAGGCGCATACAGAAATGATGTTGTTGAGTGTGAGTGTGAGTGTGAGTGTGAG 600  
QY 601 GCCCAGATTTCAATGCCCACTTGTCCGGCCCTATGTTGGGCAATTCATCTGCGCATTTGCTCT 660  
DB 601 GCCCAGATTTCAATGCCCACTTGTCCGGCCCTATGTTGGGCAATTCATCTGCGCATTTGCTCT 660  
QY 661 GACTGGAAGAACCGCTGTGCCATGCTGAAGGACGGGACTGCAAGGCTCGCACTTTCATGCGC 720  
DB 661 GACTGGAAGAACCGCTGTGCCATGCTGAAGGACGGGACTGCAAGGCTCGCACTTTCATGCGC 720  
QY 721 TCTCCGAGTGTGTGGCTTACAGCGCAGCAGCGGCCCTCACGATGACTATGTGC 780  
DB 721 TCTCCGAGTGTGTGGCTTACAGCGCAGCAGCGGCCCTCACGATGACTATGTGC 780  
QY 781 CTGCTGACCTGAAGGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGTGTGAGTGTGAG 840  
DB 781 CTGCTGACCTGAAGGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGTGTGAGTGTGAG 840  
QY 841 CAGTCGGTCTACGTTGCTACTGATTCGAGAGTTATGTGCTGAGCTCCAAACAGCTCTTC 900  
DB 841 CAGTCGGTCTACGTTGCTACTGATTCGAGAGTTATGTGCTGAGCTCCAAACAGCTCTTC 900  
QY 901 AAAGGAAGGTGAAGTGTGAGCTTGAAGCCTGAGGTGGCCAGCTGATGATC 960  
DB 901 AAAGGAAGGTGAAGTGTGAGCTTGAAGCCTGAGGTGGCCAGCTGATGATC 960  
QY 961 CTGCGCAGCGGACCACTTTATTTGGCAACTGTGTCTCCTCTTCACTGCTTGTGAAG 1020  
DB 961 CTGCGCAGCGGACCACTTTATTTGGCAACTGTGTCTCCTCTTCACTGCTTGTGAAG 1020  
QY 1021 CGGAGCGGACCTCCAGGGAGGGCGTCTTCTTTCTCGGATGAGCAGGCGCCCTAAG 1080  
DB 1021 CGGAGCGGACCTCCAGGGAGGGCGTCTTCTTTCTCGGATGAGCAGGCGCCCTAAG 1080  
QY 1081 CTGCGGACGAGTTCGTATT 1100  
DB 1081 CTGCGGACGAGTTCGTATT 1100

RESULT 4  
BD103188  
LOCUS O-fucosyltransferase.  
DEFINITION BD103188  
ACCESSION BD103188  
VERSION BD103188.1 GI:22648762  
KEYWORDS JP 2001527389-A/1.  
SOURCE unidentified  
ORGANISM unidentified  
REFERENCE 1 (bases 1 to 1514)  
AUTHORS Wang, Y. and Spellman, M.W.  
TITLE O-fucosyltransferase  
JOURNAL Patent: JP 2001527389-A 1 25-DEC-2001;  
COMMENT GENENTECH INC  
OS Unidentified



Db 436 GCATACGCTTTGAGGTGGCAGCCAGCGAAGCCAGATAAGAAAGAGCTGCCCCATGAAG 495  
Qy 361 GAAGGAAACCCCTTTGGGCCATTTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420  
Db 496 GAAGGAAACCCCTTTGGGCCATTTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 555  
Qy 421 CTTTTTACAGGATTTCTTTCAGTGTCTTCTACAGAAACAATGAGGACGAGATTTTCT 480  
Db 556 CTTTTTACAGGATTTCTTTCAGTGTCTTCTACAGAAACAATGAGGACGAGATTTTCT 615  
Qy 481 CCAAGGAACATCCGGTCTTCCCTGCCAGGAGCCAGCCAGTTCCTCCCTCTAGAA 540  
Db 616 CCAAGGAACATCCGGTCTTCCCTGCCAGGAGCCAGCCAGTTCCTCCCTCTAGAG 675  
Qy 541 GAACACAGCCCACTACAGAAATACATGTGTATGCTACAGCAAAATGCTGAAGACGGAGAG 600  
Db 676 GAACACAGCCCACTACAGAAATACATGTGTATGCTACAGCAAAATGCTGAAGACGGAGAG 735  
Qy 601 GCCCAGATTCGCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTTGCTCT 660  
Db 736 GCCCAGATTCGCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTTGCTCT 795  
Qy 661 GACTCGAAGAACCTGTGCTCATCTGAAGACGGACTGACGCTCGCACTTCATGGCC 720  
Db 796 GACTCGAAGAACCTGTGCTCATCTGAAGACGGACTGACGCTCGCACTTCATGGCC 855  
Qy 721 TCTCCGACGTGTGGGTACAGCGCAGCAGCGGCCCTCCACGATGACTATGTGC 780  
Db 856 TCTCCGACGTGTGGGTACAGCGCAGCAGCGGCCCTCCACGATGACTATGTGC 915  
Qy 781 CTGCTCAGCTGAAGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGGTCTGCGATGCC 840  
Db 916 CTGCTCAGCTGAAGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGGTCTGCGATGCC 975  
Qy 841 CAGTCGGTCTAGTGTCTACTGATTCGAGAGTTATGTGCTGAGGTCTCAACAGCTCTTC 900  
Db 976 CAGTCGGTCTAGTGTCTACTGATTCGAGAGTTATGTGCTGAGGTCTCAACAGCTCTTC 1035  
Qy 901 AAAGGAAGGTGAAGGTGTGAGCTGAAGCCTGAGGTGGCCAGCTCGACCTGTACATC 960  
Db 1036 AAAGGAAGGTGAAGGTGTGAGCTGAAGCCTGAGGTGGCCAGCTCGACCTGTACATC 1095  
Qy 961 CTGCGCCAAAGCCGACCACTTTATTTGGCAACTGTGTCTCTCTCACTGCTTTGTGAAG 1020  
Db 1096 CTGCGCCAAAGCCGACCACTTTATTTGGCAACTGTGTCTCTCTCACTGCTTTGTGAAG 1155  
Qy 1021 CCGGAGCGGACCTCGAGGAGGCGCTTCTTCTTCGCGATGAGACAGGCCCTTAAG 1080  
Db 1156 CCGGAGCGGACCTCGAGGAGGCGCTTCTTCTTCGCGATGAGACAGGCCCTTAAG 1215  
Qy 1081 CTGCGGACGAGTCTTGATT 1100  
Db 1216 CTGCGGACGAGTCTTGATT 1235

RESULT 6  
BD103189  
LOCUS BD103189 1300 bp DNA linear PAT 27-AUG-2002  
DEFINITION O-fucosyltransferase.  
ACCESSION BD103189  
VERSION BD103189.1 GI:22648763  
KEYWORDS Patent: JP 2001527389-A/2.  
SOURCE unidentified  
ORGANISM unidentified  
REFERENCE 1 (bases 1 to 1300)  
AUTHORS Wang, Y. and Spellman, M.W.  
TITLE O-fucosyltransferase  
JOURNAL Patent: JP 2001527389-A 2 25-DEC-2001;  
GENENTECH INC  
COMMENT OS Unidentified  
PN JP 2001527389-A/2  
PD 25-DEC-2001

PF 17-DEC-1997 JP 1998532877  
PR 31-JAN-1997 US 08/792498, 26-NOV-1997 US 08/978741 P1  
YANG WANG, MICHAEL W. SPELLMAN  
PC C12N15/54, C12N9/10, C07K16/40  
CC Strandedness: Single;  
Topology: Linear;  
CC O-fucosyltransferase  
FH Key Location/Qualifiers  
FT source 1..1300  
FEATURES  
source  
Location/Qualifiers  
1..1300  
/organism="unidentified"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32644"  
ORIGIN  
Query Match 99.9%; Score 1098.4; DB 6; Length 1300;  
Best Local Similarity 99.9%; Pred. No. 2.4e-253;  
Matches 1099; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 ATGCCCGCGGGCTCTGGGACCCGCGGTACCTGCTTACTGCTCCCTGCAATGGGCGC 60  
Db 136 ATGCCCGCGGGCTCTGGGACCCGCGGTACCTGCTTACTGCTCCCTGCAATGGGCGC 195  
Qy 61 TTTGGGAACACAGGCCGATCATTCTTTGGGCTCTCTGGCATTTGCAAAAGCTGCTAAACCGT 120  
Db 196 TTTGGGAACACAGGCCGATCATTCTTTGGGCTCTCTGGCATTTGCAAAAGCTGCTAAACCGT 255  
Qy 121 ACCTGGGTGCTCCCTCTTGGATTGAGTACCAAGCATCAAGCCCTCTCTTTCACCAACCTC 180  
Db 256 ACCTGGGTGCTCCCTCTTGGATTGAGTACCAAGCATCAAGCCCTCTCTTTCACCAACCTC 315  
Qy 181 CATGTGCTTACCAGAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGTCTATC 240  
Db 316 CATGTGCTTACCAGAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGTCTATC 375  
Qy 241 AGCTTGGAGGATTTTCATGGAGAGCTGGCCACCACCTGCTGCTCCCTGAGAGCGGGTG 300  
Db 376 AGCTTGGAGGATTTTCATGGAGAGCTGGCCACCACCTGCTGCTCCCTGAGAGCGGGTG 435  
Qy 301 GCATCTGCTTTGAGGTGGCAGCCAGCAAGCCAGATTAAGAAAGCTGCTCCCTCATGAAG 360  
Db 436 GCATCTGCTTTGAGGTGGCAGCCAGCAAGCCAGATTAAGAAAGCTGCTCCCTCATGAAG 495  
Qy 361 GAAGGAAACCCCTTTGGGCCATTTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420  
Db 496 GAAGGAAACCCCTTTGGGCCATTTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 555  
Qy 421 CTTTTTACAGGATTTCTTTCAGTGTCTTCTACAGAAACAATGAGGACGAGATTTTCT 480  
Db 556 CTTTTTACAGGATTTCTTTCAGTGTCTTCTACAGAAACAATGAGGACGAGATTTTCT 615  
Qy 481 CCAAGGAACATCCGGTCTTCCCTGCCAGGAGCCAGCCAGTTCCTCCCTCTAGAA 540  
Db 616 CCAAGGAACATCCGGTCTTCCCTGCCAGGAGCCAGCCAGTTCCTCCCTCTAGAG 675  
Qy 541 GAACACAGCCCACTACAGAAATACATGTGTATGCTACAGCAAAATGCTGAAGACGGAGAG 600  
Db 676 GAACACAGCCCACTACAGAAATACATGTGTATGCTACAGCAAAATGCTGAAGACGGAGAG 735  
Qy 601 GCCCAGATTCGCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTTGCTCT 660  
Db 736 GCCCAGATTCGCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTTGCTCT 795  
Qy 661 GACTCGAAGAACCTGTGCTCATCTGAAGACGGACTGACGCTCGCACTTCATGGCC 720  
Db 796 GACTCGAAGAACCTGTGCTCATCTGAAGACGGACTGACGCTCGCACTTCATGGCC 855  
Qy 721 TCTCCGACGTGTGGGTACAGCGCAGCAGCGGCCCTCCACGATGACTATGTGC 780  
Db 856 TCTCCGACGTGTGGGTACAGCGCAGCAGCGGCCCTCCACGATGACTATGTGC 915





The coding sequences of 40 new genes (K1AA0161-K1AA0200) deduced by analysis of cDNA clones from human cell line KG-1

JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS

96281124  
8724849

REFERENCE  
AUTHORS

Chiang, P.W., Wang, S., Smith, P., Song, W.J., Ramamoorthy, S., Hillman, J., Puett, S., Van Keuren, M.L., Crombez, E., Kumar, A., Glover, T.W., Miller, D.E., Tsai, C.H., Blackburn, C.C., Chen, X.N., Sun, Z., Cheng, J.F., Kornberg, J.R., and Kurnit, D.M.  
Identification and analysis of the human and murine putative chromatin structure regulator SUP6H and SUP6H

JOURNAL  
MEDLINE  
PUBMED

96374824  
8786132

REFERENCE  
AUTHORS

Ohara, O., Nagase, T., Kikuno, R., and Nomura, N.  
Direct Submission  
Submitted (12-DEC-1995) Osamu Ohara, Kazusa DNA Research Institute; 1532-3, Yana, Kisarazu, Chiba 292-0812, Japan  
(E-mail: cdna@kazusa.or.jp, Tel: +81-438-52-3913)

COMMENT  
FEATURES

On May 9, 2002 this sequence version replaced gi:1136419.  
Location/Qualifiers  
1..5189  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="ha02567s1"  
/sex="male"  
/cell\_line="KG-1"  
/cell\_type="myeloblast"  
/tissue\_type="brain"  
/clone\_lib="pBluescriptII SK plus"  
/note="This sequence was obtained by subcloning of the DNA fragments derived from two cDNA clones (1..1066 was derived from p100933 and 1067..5189 was derived from ha02567)."  
1..5189  
/gene="K1AA0180"  
1..1213  
/gene="K1AA0180"  
/citation={2}  
/codon\_start=2  
/protein\_id="BAA11497.2"  
/db\_xref="GI:20521838"  
/translation="PFRCAAGSGRADMGAAWAPLSVFLLLPLFGMPAGSW DPAGLLYCPGMRFGNQADHFLGLAFAPKLLNRTLAVPWPWIEYOHKPPPTNLHVS QKYPFLPQLQVHRVTSLEDPMELAPTHWPPEKRVAYCFEVAQSPDKKTCPMKEG NFGPFPDQPHVSFNKSELTFCIGFSASREOWSOFSPKHPHVLALPGAPQPPVLE SHRPLQKVMYSDVMYKTSQIHAHLVRPVVGLHRLIGSDWKNACAMLDGTSVGHF MASPCQGVYSRSTPAELMTMCLPDLEIORAVKLVRSLDAQSVYVATDSYSVPEL QQLFKGVKVSLLKPEVAQVDLYILGQADHFIGNCVSFTAFVKRERDLQGRPSFPG MDRPPLRLDEF"

gene  
CDS

296 CATGTGCTCTACAGAAAGTACTTCAAGCTGGAGCCCTCAGGCTTACCATCGGGTCAATC 355  
241 AGCTTGGAGGATTTTCATGGAGAAGCTGGCAACCCACCTGGCCCTCGAGAAGCGGTG 300  
356 AGCTTGGAGGATTTTCATGGAGAAGCTGGCAACCCACCTGGCCCTCGAGAAGCGGTG 415  
301 GCATATGCTTTGAGGTGGCAGCCAGCGAAGCCACAGTAAAGAACGTCGCCCATGAAG 360  
416 GCATATGCTTTGAGGTGGCAGCCAGCGAAGCCACAGTAAAGAACGTCGCCCATGAAG 475  
361 GAAGGAACCCCTTTGGCCCATTTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGAG 420  
476 GAAGGAACCCCTTTGGCCCATTTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGAG 535  
421 CTTTTCACAGGCATTTCTTCTAGTCTTCTACAGAAACAATGGAGCCAGAGATTTCT 480  
536 CTTTTCACAGGCATTTCTTCTAGTCTTCTTCTACAGAAACAATGGAGCCAGAGATTTCT 595  
481 CCAAGGAACATCGGTGCTTGGCCCTGCGAGAGCCCGCCAGCTTCCCTCTCTAGAA 540  
596 CCAAGGAACATCGGTGCTTGGCCCTGCGAGAGCCCGCCAGCTTCCCTCTCTAGAG 655  
541 GAACACAGGCCACTTACAGAAAGTACATGATGGTTCAGACGAAATGGTGAAGACGGAGAG 600  
656 GAACACAGGCCACTTACAGAAAGTACATGATGGTTCAGACGAAATGGTGAAGACGGAGAG 715  
601 GCCCAGATTCATGCCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGGGCATTTGGCTCT 660  
716 GCCCAGATTCATGCCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGGGCATTTGGCTCT 775  
661 GACTGGAAGAACCGCTTGTCCATGCTGAAGGACGGGACTCGAGGCTCGCACCTTCATGGCC 720  
776 GACTGGAAGAACCGCTTGTCCATGCTGAAGGACGGGACTCGAGGCTCGCACCTTCATGGCC 835  
721 TCTCCGCGAGTGTGGGCTACAGCGCGAGCAGCGGCCCCCTCAGCATGACTATGTC 780  
836 TCTCCGCGAGTGTGGGCTACAGCGCGAGCAGCGGCCCCCTCAGCATGACTATGTC 895  
781 CTGCTGACCTGAAGAGGATCAGAGGCGCTGTGAGCTCTGGGTGAGGTGGTGGATGCC 840  
896 CTGCTGACCTGAAGAGGATCAGAGGCGCTGTGAGCTCTGGGTGAGGTGGTGGATGCC 955  
841 CAGTCCGCTACAGTGTGCTACTGATTCGAGAGGTTATGTGCTGAGCTCCACAGCTCTTC 900  
956 CAGTCCGCTACAGTGTGCTACTGATTCGAGAGGTTATGTGCTGAGCTCCACAGCTCTTC 1015  
901 AAAGGGAAGGTGAAGGTGGTGGCTTGAAGCCTGAGGTGGCCCGCCAGCTCGACCTGTACATC 960  
1016 AAAGGGAAGGTGAAGGTGGTGGCTTGAAGCCTGAGGTGGCCCGCCAGCTCGACCTGTACATC 1075  
961 CTCGCCAAGCGCGACCACTTTATTGGCAACTGTGTCTCTCTTCACTGCTTTGTGAAG 1020  
1076 CTCGCCAAGCGCGACCACTTTATTGGCAACTGTGTCTCTCTTCACTGCTTTGTGAAG 1135  
1021 CGGAGCGGAGACCTCCAGGGGAGCGCTCTTCTTTCGCGCATGGACAGGCCCTTAAG 1080  
1136 CGGAGCGGAGACCTCCAGGGGAGCGCTCTTCTTTCGCGCATGGACAGGCCCTTAAG 1195  
1081 CTGCGGAGCGAGTTCTGATT 1100  
1196 CTGCGGAGCGAGTTCTGATT 1215

ORIGIN

Query Match 99.78; Score 1096.8; DB 9; Length 5189;  
Best Local Similarity 99.84; Pred. No. 5.9e-253; Indels 0; Gaps 0;  
Matches 1098; Conservative 0; Mismatches 2;

Qy 1 ATGCCCGGGCTCTGGGACCGCGCGGTACTGCTCTACTGCGCCCTGCGATGGGGCGC 60  
Db 116 ATGCCCGGGCTCTGGGACCGCGCGGTACTGCTCTACTGCGCCCTGCGATGGGGCGC 175  
Qy 61 TTGGGAACAGGCGCATCTTTGGGCTCTCTGGGATTTGCAAGCTGTCAACCGT 120  
Db 176 TTGGGAACAGGCGCATCTTTGGGCTCTCTGGGATTTGCAAGCTGTCAACCGT 235  
Qy 121 ACCTTGGGTGCTCCCTCTGGATGAGTACAGGATCACAAGCTCTCTTCCACCAACCTC 180  
Db 236 ACCTTGGGTGCTCCCTCTGGATGAGTACAGGATCACAAGCTCTCTTCCACCAACCTC 295  
Qy 181 CATGTGCTCTACAGAAAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGCTCATC 240

RESULT 10  
CQ727777

LOCUS 1167 bp DNA linear PAT 03-FEB-2004  
DEFINITION Sequence 13711 from Patent WO02068579.  
ACCESSION CQ727777  
VERSION CQ727777.1 GI:42294771  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Db	850	CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCTGCTGGATGCC	909			
Qy	841	CAGTCGGTCTACGTTGCTACTGATTCGAGAGTTATGTGCTCAGCTCCAAACAGCTCTTC	900			
Db	910	CAGTCGGTCTACGTTGCTACTGATTCGAGAGTTATGTGCTCAGCTCCAAACAGCTCTTC	969			
Qy	901	AAAGGGAAGGTGAAGGTGTTGAGCCTGAAGCCTGAAGGTGGCCAGGTGCACTGTATCATC	960			
Db	970	AAAGGGAAGGTGAAGGTGTTGAGCCTGAAGCCTGAAGGTGGCCAGGTGCACTGTATCATC	1029			
Qy	961	CTGGCCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCTTCACTGCCTTTGTGAAG	1020			
Db	1030	CTGGCCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCTTCACTGCCTTTGTGAAG	1089			
Qy	1021	CGGGAGCGGACCTCCAGGGGAGCGCTCTTCTTTCTTCGGCATGGACAGGCCCTTAAG	1080			
Db	1090	CGGGAGCGGACCTCCAGGGGAGCGCTCTTCTTTCTTCGGCATGGACAGGCCCTTAAG	1149			
Qy	1081	CTGGGGACGAGTTCTGA	1098			
Db	1150	CTGGGGACGAGTTCTGA	1167			
RESULT 11						
AX780153						
LOCUS	AX780153	5266 bp	DNA			
DEFINITION	Sequence 2310 from Patent WO03039443.					
ACCESSION	AX780153					
VERSION	AX780153.1					
KEYWORDS	GI:32697147					
SOURCE	Homo sapiens (human)					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
AUTHORS	1. Haerlath, T., Schoch, C., Kern, W., Kohlmann, A., Schnittger, S., Dugas, M., Bils, R., Brors, B. and Mergenthaler, S.					
TITLE	Novel genetic markers for leukemias					
JOURNAL	Patent: WO 03039443-A 2310 15-MAY-2003; Deutsches Krebsforschungszentrum (DK) ; Ludwig-Maximilian-Universitaet Muenchen (DE) ; Haerlath, Torsten, PD Dr. Dr. (DE) ; Schoch, Claudia (DE) ; Kern, Wolfgang (DE)					
FEATURES	Location/Qualifiers					
source	1..5266					
/organism="Homo sapiens"						
/mol_type="unassigned DNA"						
/db_xref="taxon:9606"						
ORIGIN						
Query Match 99.6%; Score 1095.8; DB 6; Length 5266;						
Best Local Similarity 99.7%; Pred. No. 1e-252; 3; Indels 0; Gaps 0;						
Matches 1097; Conservative 0; Mismatches 3; Indels 0; Gaps 0;						
Qy	1	ATGCCCGCGGGCTCCCTGGACCGCGCGGTACCTGCTCTACTGCGCCCTGCATGGGGCGC	60			
Db	132	ATGCTCTGGGGCTCCCTGGACCGCGCGGTACCTGCTCTACTGCGCCCTGCATGGGGCGC	191			
Qy	61	TTTGGGAAACAGGCCCGATCACTTTCTTGGGCTCTCTGGCATTTGCAAAAGCTGTAAACCGT	120			
Db	192	TTTGGGAAACAGGCCCGATCACTTTCTTGGGCTCTCTGGCATTTGCAAAAGCTGTAAACCGT	251			
Qy	121	ACCTTGGCTGCTCCCTCTTGGATTGATGACAGCATCAAGCCCTCTTTCACCACTC	180			
Db	252	ACCTTGGCTGCTCCCTCTTGGATTGATGACAGCATCAAGCCCTCTTTCACCACTC	311			
Qy	181	CATGTTCTCTACAGAAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGGTCATC	240			
Db	312	CATGTTCTCTACAGAAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGGTCATC	371			
Qy	241	AGCTTGGAGATTTCATGAGAACTGGACCCACCACCTCTGGCCCTCTGAGAGCGGGTG	300			
Db	372	AGCTTGGAGATTTCATGAGAACTGGACCCACCACCTCTGGCCCTCTGAGAGCGGGTG	431			



Qy	301	GCATACCTGCTTTGAGTGGAGCCGAGCCAGGAGCCAGATGAAGAGACGTGCCCCATGAAG	360
Db	432	GCATACCTGCTTTGAGTGGAGCCGAGCCAGGAGCCAGATGAAGAGACGTGCCCCATGAAG	491
Qy	361	GAAGAAACCCCTTTGGCCCATCTTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG	420
Db	492	GAAGAAACCCCTTTGGCCCATCTTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG	551
Qy	421	CTTTTTCACAGCATTTCTTTCAGTGTCTTCTACAGAAACATGAAGCCAGAGATTTTCT	480
Db	552	CTTTTTCACAGCATTTCTTTCAGTGTCTTCTACAGAAACATGAAGCCAGAGATTTTCT	611
Qy	481	CCAAAGGAACATCGGTGCTTGGCTGCGAGGACCCAGCCAGTCCCGTCTTAGAA	540
Db	612	CCAAAGGAACATCGGTGCTTGGCTGCGAGGACCCAGCCAGTCCCGTCTTAGAG	671
Qy	541	GAACACAGGCGACTACAGAAATACATGTATGGTCAGACGAAATGGTGAAGACGGAGAG	600
Db	672	GAACACAGGCGACTACAGAAATACATGTATGGTCAGACGAAATGGTGAAGACGGAGAG	731
Qy	601	GCCAGATTCATGCCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTTGGCTCT	660
Db	732	GCCAGATTCATGCCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTTGGCTCT	791
Qy	661	GACTGGAGAACGCTGTGCGATGCTGAAGACGGGACTCGAGGCTCGCACTTCATGGCC	720
Db	792	GACTGGAGAACGCTGTGCGATGCTGAAGACGGGACTCGAGGCTCGCACTTCATGGCC	851
Qy	721	TCTCCGAGTGTGTGGCTTACAGCGCGAGACAGCGGCCCCCTTACGATGACTATGTGC	780
Db	852	TCTCCGAGTGTGTGGCTTACAGCGCGAGACAGCGGCCCCCTTACGATGACTATGTGC	911
Qy	781	CTGCTGACCTTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCTGCGATGCC	840
Db	912	CTGCTGACCTTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCTGCGATGCC	971
Qy	841	CAGTCGGTCTACGTGCTACTGATTCAGAGAGTTATGTGCTCAGCTCCAAACAGCTCTTC	900
Db	972	CAGTCGGTCTACGTGCTACTGATTCAGAGAGTTATGTGCTCAGCTCCAAACAGCTCTTC	1031
Qy	901	AAAGGGAAGTGAAGTGTGAGCTTGAAGCTGAGGTGCGCCAGCTGACCTGTACATC	960
Db	1032	AAAGGGAAGTGAAGTGTGAGCTTGAAGCTGAGGTGCGCCAGCTGACCTGTACATC	1091
Qy	961	CTCGGCCAAGCCGACCACTTTATTTGGCAACTGTGTCTCTCTTCACTGCCCTTTGTGAAG	1020
Db	1092	CTCGGNCAAGCCGACCACTTTATTTGGCAACTGTGTCTCTCTTCACTGCCCTTTGTGAAG	1151
Qy	1021	CGGGAGCGGGACCTCCAGGGGAGCGCTCTTCTTTCTTGGGCATGACAGGCCCCCTTAAG	1080
Db	1152	CGGGAGCGGGACCTCCAGGGGAGCGCTCTTCTTTCTTGGGCATGACAGGCCCCCTTAAG	1211
Qy	1081	CTGCGGACGAGTTCTGATT	1100
Db	1212	CTGCGGACGAGTTCTGATT	1231
RESULT 12			
HSM803330			
LOCUS			
DEFINITION Homo sapiens mRNA; cDNA DKFp451J114 (from clone DKFp451J114).			
ACCESSION AL832023			
VERSION AL832023.1			
GI:21732563			
KEYWORDS			
SOURCE Homo sapiens (human)			
ORGANISM			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
REFERENCE 1 (bases 1 to 4560)			
AUTHORS Wambutt,R., Heubner,D., Mewes,H.W., Weil,B., Amid,C., Osanger,A.,			
Fobo,G., Han,W. and Wiemann,S.			
TITLE Direct Submission			
JOURNAL Submitted (13-MAY-2003) MIPS, Ingolstaedter Landstr.1, D-85764			

COMMENT	Neuberberg, GERMANY Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing consortium of the German Genome Project. This clone (DKFp451J114) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzd.de Further information about the clone and the sequencing project is available at http://mips.gsf.de/proj/cDNA/.			
FEATURES	Location/Qualifiers			
source	1..4560			
	/organism="Homo sapiens"			
	/mol_type="mRNA"			
	/db_xref="RZPD:DKFp451J114"			
	/db_xref="taxon:9606"			
	/clone="DKFp451J114"			
	/tissue_type="human skeletal muscle"			
	/clone_lib="451 (synonym: hlcc1). Vector pSport1; host DH10B; sites NotI + SalI"			
	/dev_stage="adult"			
	4538			
polyA_site				
ORIGIN				
Query Match	99.6%; Score 1095.2; DB 9; Length 4560;			
Best Local Similarity	99.7%; Pred. No. 1.4e-252;			
Matches 1097; Conservative	0; Mismatches 3; Indels 0; Gaps 0;			
Qy	1	ATGCCCGGGCTCTCTGGGACCGCGCGGTTACCTGCTCTACTGCCCCCTGCATGGGGCGC	60	
Db	51	ATGCCCTGGGGCTCTCTGGGACCGCGCGGTTACCTGCTCTACTGCCCCCTGCATGGGGCGC	110	
Qy	61	TTTGGGAACCAAGCCGATCATCTTCTTGGGCTCTCTGGCATTTGCAAGCTGTAACCGGT	120	
Db	111	TTTGGGAACCAAGCCGATCATCTTCTTGGGCTCTCTGGCATTTGCAAGCTGTAACCGGT	170	
Qy	121	ACCTTGGCTGTCTCTCTGGATTTGATACCAAGCTCACAAGCTCTCTTTCACCAACCTC	180	
Db	171	ACCTTGGCTGTCTCTCTGGATTTGATACCAAGCTCACAAGCTCTCTTTCACCAACCTC	230	
Qy	181	CATGTGCTCTACAGAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGCTCATC	240	
Db	231	CATGTGCTCTACAGAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGCTCATC	290	
Qy	241	AGCTTGGAGGATTTTCAATGGAGAGCTGGCACTCCACCTGCGCCCTCTGAGAAGCGGTG	300	
Db	291	AGCTTGGAGGATTTTCAATGGAGAGCTGGCACTCCACCTGCGCCCTCTGAGAAGCGGTG	350	
Qy	301	GCATACTCTTCTGAGGTGGCAGCCAGCGAAGCCAGATAAGAGACGTGCCCATGAAG	360	
Db	351	GCATACTCTTCTGAGGTGGCAGCCAGCGAAGCCAGATAAGAGACGTGCCCATGAAG	410	
Qy	361	GAAGGAACCCCTTTGGCCCATTTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG	420	
Db	411	GAAGGAACCCCTTTGGCCCATTTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG	470	
Qy	421	CTTTTTCAGGCAATTTCTTCAGTGTCTTCTACAGAGCAATGGAGCCAGAGATTTTCT	480	
Db	471	CTTTTTCAGGCAATTTCTTCAGTGTCTTCTACAGAGCAATGGAGCCAGAGATTTTCT	530	
Qy	481	CCAAAGGAACATCCGCTGTCTTCTTCCCTGCGAGAGCCCGAGCCAGATTTCCCGCTCTAGAA	540	
Db	531	CCAAAGGAACATCCGCTGTCTTCTTCCCTGCGAGAGCCCGAGCCAGATTTCCCGCTCTAGAG	590	
Qy	541	GAACACAGCCCACTACAGAAATGATGTTATGTTATGTTATGTTATGTTATGTTATGTTATGTT	600	
Db	591	GAACACAGCCCACTACAGAAATGATGTTATGTTATGTTATGTTATGTTATGTTATGTTATGTT	650	
Qy	601	GCCAGATTCATGCCACCTTGTGCGGCTTATGTGGGCAATTCATCTGCGCATTTGGCTCT	660	
Db	651	GCCAGATTCATGCCACCTTGTGCGGCTTATGTGGGCAATTCATCTGCGCATTTGGCTCT	710	
Qy	661	GACTGGAAGAACGGCTGTGSCCATCTCTGAAGAGACGGGACTGCAAGCTCGCATTCATGGCC	720	



Db	711	GACTGGAGAACGCTGTGCCATCTCTGAAGACGGGACTGGAGCTCGCACTTCATGGCC	770
Qy	721	TCTCCGCAGTGTGTGGGCTACAGCCGCAGACACAGCGGCCCCCTTCACGATGACTATGTGC	780
Db	771	TCTCCGCAGTGTGTGGGCTACAGCCGCAGACACAGCGGCCCCCTTCACGATGACTATGTGC	830
Qy	781	CTGCCTGACCTGAAGAGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCCGTGGATGCC	840
Db	831	CTGCCTGACCTGAAGAGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCCGTGGATGCC	890
Qy	841	CAGTCGGTCTACGTTGTCTACTGATTCGAGAGTTATGTGCGCTGAGCTCCAAAGCTCTTC	900
Db	891	CAGTCGGTCTACGTTGTCTACTGATTCGAGAGTTATGTGCGCTGAGCTCCAAAGCTCTTC	950
Qy	901	AAAGGGAAGTGAAGGTGTGTGAGCCTCAAGCTGAGGTGGCCCAAGTCGACCTGTATCATC	960
Db	951	AAAGGGAAGTGAAGGTGTGTGAGCCTCAAGCTGAGGTGGCCCAAGTCGACCTGTATCATC	1010
Qy	961	CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCTCTTCCTGCTGCTTTGTGAAG	1020
Db	1011	CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCTCTTCCTGCTGCTTTGTGAAG	1070
Qy	1021	CGGAGCGGACCTCCAGGGAGGACCGCTCTTCTTTCTTCGGCATGGACAGGCCCCCTTAAG	1080
Db	1071	CGGAGCGGACCTCCAGGGAGGACCGCTCTTCTTTCTTCGGCATGGACAGGCCCCCTTAAG	1130
Qy	1081	CTCGGGACAGATTCTGATT	1100
Db	1131	CTCGGGACAGATTCTGATT	1150

```

RESULT 13
AJ781500
LOCUS      1167 bp      mRNA      linear      PRI 07-JUL-2004
DEFINITION Pan troglodytes mRNA for protein-O-fucosyltransferase 1 (fut12 gene).
ACCESSION AJ781500
VERSION    AJ781500.1 GI:50057069
KEYWORDS   fut12 gene; pofut1 gene; protein-O-fucosyltransferase 1.
SOURCE     Pan troglodytes (chimpanzee)
ORGANISM   Pan troglodytes
            Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
REFERENCE 1 Martinez-Duncker, I., Mollicone, R., Candelier, J.J., Breton, C. and Oriol, R.
            A new superfamily of protein-O-fucosyltransferases,
            alpha2-fucosyltransferases, and alpha6-fucosyltransferases:
            phylogeny and identification of conserved peptide motifs
            Glycobiology 13 (12), 1C-5C (2003)
            12966037
REFERENCE 2 Martinez-Duncker, I., Oriol, R. and Mollicone, R.
            Phylogeny of fucosyltransferases
            Unpublished
            3 (bases 1 to 1167)
            Oriol, R.
            Direct Submission
            Submitted (06-JUL-2004) Oriol R., U504, Inserm, 16 Av. Paul
            Vaillant-Couturier, 94807, FRANCE
FEATURES             source
             1..1167
                /organism="Pan troglodytes"
                /mol_type="mRNA"
                /db_xref="taxon:9598"
             1..1167
                /gene="fut12"
             1..1167
                /gene="fut12"
                /function="adds fucose to Ser or Thr"
                /note="pofut1"
                /codon_start=1

```



source

1..5009

/organism="unidentified"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32644"

ORIGIN

Query Match 93.9%; Score 1033.4; DB 6; Length 5009;

Best Local Similarity 99.9%; Pred. No. 1e-237;

Matches 1034; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	66	GAACACAGCCGATCACTTCTGGGCTCTGGCATTTGCAAGCTGTAAACCGTACCTT	125
DB	1	GAACACAGCCGATCACTTCTGGGCTCTGGCATTTGCAAGCTGTAAACCGTACCTT	60
QY	126	GGCTGTCCCTCTGGATTGATACAGCATCACAAAGCTCTTTTCAACCACTCCATGT	185
DB	61	GGCTGTCCCTCTGGATTGATACAGCATCACAAAGCTCTTTTCAACCACTCCATGT	120
QY	186	GTCTTACAGAAAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGGTCATCAGCTT	245
DB	121	GTCTTACAGAAAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGGTCATCAGCTT	180
QY	246	GGAGGATTTCTAGGAGAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTGGCATA	305
DB	181	GGAGGATTTCTAGGAGAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTGGCATA	240
QY	306	CTGCTTTGAGGTGGCAGCCAGCGAAGCCAGATAAGAAAGCTGCCCCCATGAAGGAAGG	365
DB	241	CTGCTTTGAGGTGGCAGCCAGCGAAGCCAGATAAGAAAGCTGCCCCCATGAAGGAAGG	300
QY	366	AAACCCCTTTGGCCCATTTCTGGGATCAGTTTCATGTGATTTCAAACAAGTCGGAGCTTTT	425
DB	301	AAACCCCTTTGGCCCATTTCTGGGATCAGTTTCATGTGATTTCAAACAAGTCGGAGCTTTT	360
QY	426	TACAGGCATTTCTTTCAGTGCTTCTACAGAGAACTTGGAGCCAGAGATTTTCTCCAAA	485
DB	361	TACAGGCATTTCTTTCAGTGCTTCTACAGAGAACTTGGAGCCAGAGATTTTCTCCAAA	420
QY	486	GGAACATCCGGTGTCTTGGCCCTGCCAGGAGCCCGCCAGTTTCCCGCTCTAGAGAAACA	545
DB	421	GGAACATCCGGTGTCTTGGCCCTGCCAGGAGCCCGCCAGTTTCCCGCTCTAGAGAAACA	480
QY	546	CAGGCCACTACAGAAAGTACATGGTATGTCAGACGAAATGTTGAAGCGGAGAGGCCCA	605
DB	481	CAGGCCACTACAGAAAGTACATGGTATGTCAGACGAAATGTTGAAGCGGAGAGGCCCA	540
QY	606	GATTTCATGCCACCTTGTCCGGCCCTATGTGGGCATTTTCATCTGGCCATTTGCTGACTG	665
DB	541	GATTTCATGCCACCTTGTCCGGCCCTATGTGGGCATTTTCATCTGGCCATTTGCTGACTG	600
QY	666	GAAGAACGCTGTGTCATGCTGAAGGACGGGACTGACGGCTCGCACTTTCATGGCCCTCTCC	725
DB	601	GAAGAACGCTGTGTCATGCTGAAGGACGGGACTGACGGCTCGCACTTTCATGGCCCTCTCC	660
QY	726	GCAGTGTGTGGGCTACAGCCGAGCAGACGCGGCCCTCCAGATGACTATGTGCTGCC	785
DB	661	GCAGTGTGTGGGCTACAGCCGAGCAGACGCGGCCCTCCAGATGACTATGTGCTGCC	720
QY	786	TGACCTGAAGAGATCAGAGGGCTGTGAAGCTCTGGGTGAGGTGCTGGATGCCAGTC	845
DB	721	TGACCTGAAGAGATCAGAGGGCTGTGAAGCTCTGGGTGAGGTGCTGGATGCCAGTC	780
QY	846	GGTCTACGCTTGTCTACTGATTTCCGAGAGTTATGTGCTGAGCTCCAAACAGCTCTTCAAAGG	905
DB	781	GGTCTACGCTTGTCTACTGATTTCCGAGAGTTATGTGCTGAGCTCCAAACAGCTCTTCAAAGG	840
QY	906	GAAGGTGAAGGTGTGAGCCTGAAGCCTGAGGTGGCCAGGTTCGACCTGTATCATCTCTCGG	965
DB	841	GAAGGTGAAGGTGTGAGCCTGAAGCCTGAGGTGGCCAGGTTCGACCTGTATCATCTCTCGG	900
QY	966	CAAAGCCGACACTTTATTTGGCACTGTCTCTCTCTTCACTGCTTTTGAAGCGGGA	1025
DB	901	CAAAGCCGACACTTTATTTGGCACTGTGTCTCTCTCTTCACTGCTTTTGAAGCGGGA	960

Search completed: November 9, 2005, 07:59:59  
Job time : 4964.62 secs

QY	1026	GGGGGACCTCCAGGGGAGGCGCTCTTCTTTCTTGGCATGGACAGGCCCTTAAGCTGG	1085
DB	961	GGGGGACCTCCAGGGGAGGCGCTCTTCTTTCTTGGCATGGACAGGCCCTTAAGCTGG	1020
QY	1086	GGACGAGTTCTGATT	1100
DB	1021	GGACGAGTTCTGATT	1035

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2005, 23:23:13 ; Search time 642.158 Seconds  
(without alignments)  
10140.369 Million cell updates/sec

Title: US-09-774-954-16

Perfect score: 1100

Sequence: 1 ATGCCCGCGGGCTCTCTGGGA.....CTGCGGACGAGTTCGATT 1100

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_16Dec04:\*

- 1: Geneseqn1980s:\*
- 2: Geneseqn1990s:\*
- 3: Geneseqn2000s:\*
- 4: Geneseqn2001s:\*
- 5: Geneseqn2001bs:\*
- 6: Geneseqn2002as:\*
- 7: Geneseqn2002bs:\*
- 8: Geneseqn2003as:\*
- 9: Geneseqn2003bs:\*
- 10: Geneseqn2003cs:\*
- 11: Geneseqn2003ds:\*
- 12: Geneseqn2004as:\*
- 13: Geneseqn2004bs:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1100	100.0	1514	2	AAV65632
2	1098.4	99.9	11284	2	AAV65633
3	1096.8	99.7	5218	13	ACN40746
4	1095.8	99.6	5266	10	ADF81754
5	1046.2	95.1	5230	4	AAK51510
6	1041.4	94.7	4850	4	AAK52494
7	1033.4	93.9	5009	2	AAV65634
8	473.4	43.0	915	4	AD07629
9	406.2	36.9	610	13	ADQ56824
10	251.2	22.8	10331	4	AAK79915
11	251.2	22.8	10331	4	AAK79915
12	233	21.2	479	9	ACH32225
13	221.8	20.2	477	9	ACH35532
14	166.4	15.1	1209	4	ABL13947
c 15	150.4	13.7	3264	4	ABL13946
c 16	123.2	11.2	1545	4	AAF23894
c 17	108.6	9.9	3793	4	ABL13894
c 18	71.8	6.5	3567	4	ABL13952
c 19	60	5.5	60	6	ABN44539
20	56	5.1	760	3	AAZ98210

21	50	4.5	2000	8	ADA71938
c 22	46.2	4.2	2000	8	ADA71938
23	42.4	3.9	1438	8	ACA31561
c 24	41.2	3.7	114793	4	AD08215
c 25	40.6	3.7	558	5	AAH97782
26	40.4	3.7	5019	13	ADQ80232
27	40.4	3.7	5704	12	ADQ84111
28	40.4	3.7	5932	10	ADB75488
29	40	3.6	40	2	AAV65637
30	40	3.6	40	2	AAV65635
31	40	3.6	1554	8	ADA69815
32	39	3.5	372	8	ACA43606
33	39	3.5	1131	11	ABD01472
34	39	3.5	1353	11	ABD01484
35	38.6	3.5	407	8	ABX46567
36	38.6	3.5	1596	10	ADI62782
37	38.6	3.5	2358	12	ADE77025
38	38.6	3.5	2670	10	ADE25660
39	38.4	3.5	532	9	ACH26305
40	38.4	3.5	6227	6	ABT10150
41	38.4	3.5	6227	8	ACA64963
c 42	37.8	3.4	1234	13	ADS63598
c 43	37.8	3.4	1234	13	ADS63974
c 44	37.8	3.4	1234	13	ADT41577
45	37.8	3.4	2370	4	AAH52044

## ALIGNMENTS

RESULT 1

AAV65632

ID AAV65632 standard; DNA; 1514 BP.

XX

AC AAV65632;

XX

DT 16-DEC-1998 (first entry)

XX

DE Human heart O-fucosyltransferase encoding DNA.

XX

XX O-fucosyltransferase; epidermal growth factor; EGF; glycosylation;

KW O-fucose; inhibitor; sensory neuron; retinal neuron; human; heart; ss.

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT misc\_feature 1..1100

FT /\*tag= b

FT /note= "this actively expressed O-fucosyltransferase

FT sequence is claimed for in claim 9"

FT 1..1098

FT /\*tag= a

FT /product= "human heart O-fucosyltransferase"

XX

XX WO9833924-A1.

XX

PD 06-AUG-1998.

XX

PF 17-DEC-1997; 97WO-US023401.

XX

PR 31-JAN-1997; 97US-00792498.

XX

PR 26-NOV-1997; 97US-00978741.

XX

XX (GETH ) GENENTECH INC.

PA Wang Y, Spellman MW;

XX

PI WPI; 1998-437477/37.

XX

DR P-PSDB; AAW80571.

XX

Human O-fucosyltransferase able to glycosylate epidermal growth factor domains - useful for diagnosis and treatment of diseases involving overexpression of the enzyme.



CC O-fucosyltransferase, e.g. mutants with increased affinity for the EFG  
CC domains, are used in diagnosis and treatment of conditions associated  
CC with overexpression of O-fucosyltransferase, to promote survival of  
CC sensory (retinal) neurons. Probes based on EGF domain polypeptide are  
CC used to detect gene amplification and expression. The expression can also  
CC be determined at the protein level using antibodies specific for O-  
CC fucosyltransferase  
XX

SQ Sequence 11284 BP; 2947 A; 2602 C; 2624 G; 3111 T; 0 U; 0 Other;

Query Match 99.9%; Score 1098.4; DB 2; Length 11284;  
Best Local Similarity 99.9%; Pred. No. 1.3e-305;  
Matches 1099; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy	1	ATGCCCGCGGGCTCTCTGGGACCCGCGGTTACCTGCTCTACTGCTCCCTGATGGGGCGC	60
Db	4236	ATGCCCGCGGGCTCTCTGGGACCCGCGGTTACCTGCTCTACTGCTCCCTGATGGGGCGC	4295
Qy	61	TTTGGGAACCAAGCGCATCACTTTCTTGGGCTCTCTGGCAATTTGCAAGCTGCTAAACCGT	120
Db	4296	TTTGGGAACCAAGCGCATCACTTTCTTGGGCTCTCTGGCAATTTGCAAGCTGCTAAACCGT	4355
Qy	121	ACCTTGGCTGCTCCCTCTTGGATTGAGTACAGCATCAAGACCTCTCTTACCAACCTC	180
Db	4356	ACCTTGGCTGCTCCCTCTTGGATTGAGTACAGCATCAAGACCTCTCTTACCAACCTC	4415
Qy	181	CATGTGCTCTACCAAGTACTTCAAGCTGAGGCCCTCCAGGCTTACCATCGGTCATC	240
Db	4416	CATGTGCTCTACCAAGTACTTCAAGCTGAGGCCCTCCAGGCTTACCATCGGTCATC	4475
Qy	241	AGCTTGGAGGATTTTCATGGAAGCTGCGACCCACCCACTGCGCCCTTGAGAAGCGGGTG	300
Db	4476	AGCTTGGAGGATTTTCATGGAAGCTGCGACCCACCCACTGCGCCCTTGAGAAGCGGGTG	4535
Qy	301	GCATACTCTTTAGTGGGAGCCAGCAAGCCAGATAGAGAGAGTGCCCATGAG	360
Db	4536	GCATACTCTTTAGTGGGAGCCAGCAAGCCAGATAGAGAGAGTGCCCATGAG	4595
Qy	361	GAAGAAACCCCTTTGGCCCACTTCGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG	420
Db	4596	GAAGAAACCCCTTTGGCCCACTTCGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG	4655
Qy	421	CTTTTTCAGGCAATTCCTTCAGTGCTTCTTACAGAGAAATGAGGCGCAGAGATTTTCT	480
Db	4656	CTTTTTCAGGCAATTCCTTCAGTGCTTCTTACAGAGAAATGAGGCGCAGAGATTTTCT	4715
Qy	481	CCAAAGGAACATCGGTGCTTCCCTGCGGAGCCCGCCAGGCTTCCCGTCTAGAA	540
Db	4716	CCAAAGGAACATCGGTGCTTCCCTGCGGAGCCCGCCAGGCTTCCCGTCTAGAG	4775
Qy	541	GAACACAGGCCACTACAGAGTACATGTATGTTGTCAGACGAAATGGTGAACGCGGAG	600
Db	4776	GAACACAGGCCACTACAGAGTACATGTATGTTGTCAGACGAAATGGTGAACGCGGAG	4835
Qy	601	GCCAGATTCATGCCACCTTGTCCGCGCCCTATGTGGCAATTCATCTGCGCATTTGGCTCT	660
Db	4836	GCCAGATTCATGCCACCTTGTCCGCGCCCTATGTGGCAATTCATCTGCGCATTTGGCTCT	4895
Qy	661	GACTGGAAGACCGCTGTGCCATGCTGAAGGACGGGACTGCGGCTGCGCATTCATGGCC	720
Db	4896	GACTGGAAGACCGCTGTGCCATGCTGAAGGACGGGACTGCGGCTGCGCATTCATGGCC	4955
Qy	721	TCTCCGAGTGTGGGCTACAGCGGACGACAGCGGCCCTCCAGATGACTATGTGC	780
Db	4956	TCTCCGAGTGTGGGCTACAGCGGACGACAGCGGCCCTCCAGATGACTATGTGC	5015
Qy	781	CTGCTGACCTGGAAGGAGATCCAGAGGCTCTGGAAGCTCTGGGTGAGGTCTGGATGCC	840
Db	5016	CTGCTGACCTGGAAGGAGATCCAGAGGCTCTGGAAGCTCTGGGTGAGGTCTGGATGCC	5075
Qy	841	CAGTCGGTCTAGTGTGCTACTGATTCGAGAGTTATGTGCTGAGCTCCAAAGCTCTTC	900
Db	5076	CAGTCGGTCTAGTGTGCTACTGATTCGAGAGTTATGTGCTGAGCTCCAAAGCTCTTC	5135

Qy	901	AAAGGGAAGTGAAGGTGGTGGAGCTGAAGCTGAGGTGGCCAGGTGCGACCTGTACATC	960
Db	5136	AAAGGGAAGTGAAGGTGGTGGAGCTGAAGCTGAGGTGGCCAGGTGCGACCTGTACATC	5195
Qy	961	CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCTCTCTTCACTGCGCTTTGTGAAG	1020
Db	5196	CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCTCTCTTCACTGCGCTTTGTGAAG	5255
Qy	1021	CGGAGGCGGACCTCCAGGAGGCGCTCTTCTTCTCGCATGGACAGGCCCTTAAG	1080
Db	5256	CGGAGGCGGACCTCCAGGAGGCGCTCTTCTTCTTCTCGCATGGACAGGCCCTTAAG	5315
Qy	1081	CTCGGGACGAGTCTTGATT	1100
Db	5316	CTCGGGACGAGTCTTGATT	5335

#### RESULT 3

ACN40746	ID	ACN40746	standard; cDNA; 5218 BP.
XX	AC	ACN40746;	
XX	DT	18-NOV-2004	(first entry)
XX	XX	Tumour-associated antigenic target (TAT)	CDNA DNA326813, SEQ ID NO:5725.
XX	DE	Tumour-associated antigenic target; TAT; human; overexpression; cancer;	
XX	KW	tumour; diagnosis; cell proliferative disorder; breast cancer;	
XX	KW	colorectal cancer; lung cancer; ovarian cancer; liver cancer;	
XX	KW	central nervous system cancer; bladder cancer; pancreatic cancer;	
XX	KW	cervical cancer; melanoma; leukaemia; hybridisation probe;	
XX	KW	chromosome identification; chromosome mapping; gene mapping;	
XX	OS	Homo sapiens.	
XX	PN	WO2004030615-A2.	
XX	XX	15-APR-2004.	
XX	PD	29-SEP-2003; 2003WO-US028547.	
XX	PF	02-OCT-2002; 2002US-0414971P.	
XX	PR	(GETH ) GENENTECH INC.	
XX	PA	Wu TD, Zhang Z, Zhou Y;	
XX	PI	WPI; 2004-347921/32.	
XX	PI	P-PSDB; ABM82223.	
XX	DR	New tumor-associated antigenic target polypeptides and nucleic acids,	
XX	DR	useful in preparing a medicament for treating or detecting a	
XX	DR	proliferative disorder, e.g. breast, lung, colorectal, ovarian or	
XX	DR	prostate cancer or tumor.	
XX	PT	Claim 1; SEQ ID NO 5725; 7273pp; English.	
XX	PS	The invention relates to human tumour-associated antigenic target (TAT)	
XX	CC	polypeptides, and their related nucleic acids. The TAT polypeptides are	
XX	CC	overexpressed in cancer tissues compared to normal tissues, and may thus	
XX	CC	serve as effective targets for the diagnosis and treatment of cancer in	
XX	CC	mammals. The invention also relates to nucleic acid and polypeptide	
XX	CC	sequences at least 80% identical to the TAT nucleic acids and	
XX	CC	polypeptides; expression vectors and host cells comprising a TAT nucleic	
XX	CC	acid; an antibody specific for a TAT polypeptide; a peptide or organic	
XX	CC	molecule which binds to a TAT polypeptide; fusion proteins comprising a	
XX	CC	TAT polypeptide; and methods and compositions for the treatment or	
XX	CC	diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,	
XX	CC	antibodies, antagonists, binding molecules and compositions are useful	
XX	CC	for diagnosing or treating a cell proliferative disorder associated with	







110 TGCTAAACCGTACTTGGCTGTCTCTCTTGGATTGATACAGCATCAAGCTCTCTT 169  
Db TGCTAAACCGTACTTGGCTGTCTCTCTTGGATTGATACAGCATCAAGCTCTCTT 325  
Qy TCACCAACCTCCATGTGTCTTACCAAGAGTACTTCAAGCTGGAGCCCTTCCAGGTTACC 229  
Db TCACCAACCTCCATGTGTCTTACCAAGAGTACTTCAAGCTGGAGCCCTTCCAGGTTACC 385  
Qy ATCCGGTTCATCAGCTTGGAGGATTTTCATGGAGAGCTGGCACCACCACTGGCCCTG 289  
Db ATCCGGTTCATCAGCTTGGAGGATTTTCATGGAGAGCTGGCACCACCACTGGCCCTG 445  
Qy AGAAGCGGTGGCATACTGCTTTGAGTGGCAGCCAGCCAGCCAGATAGAGAGCT 349  
Db AGAAGCGGTGGCATACTGCTTTGAGTGGCAGCCAGCCAGCCAGATAGAGAGCT 505  
Qy GCCCATGAAGAGGAACCCCTTTGGCCCAATCTCTGGATCAGTTTTCATGTGAGTTTCA 409  
Db GCCCATGAAGAGGAACCCCTTTGGCCCAATCTCTGGATCAGTTTTCATGTGAGTTTCA 565  
Qy ACAAGTCGGAGCTTTTACAGGATTTCTCTTCAAGTCTTCTTACAGAGAACTGGAGCC 469  
Db ACAAGTCGGAGCTTTTACAGGATTTCTCTTCAAGTCTTCTTACAGAGAACTGGAGCC 625  
Qy AGAGATTTTCTCAAAGGAACATCGGTGCTTGGCTCCAGAGCCCGAGCCAGTTCC 529  
Db AGAGATTTTCTCAAAGGAACATCGGTGCTTGGCTCCAGAGCCCGAGCCAGTTCC 685  
Qy CGCTCTAGAGAACACAGGCCATACAGAACTATGCTATGCTCAGACGAATGTGGA 589  
Db CGCTCTAGAGAACACAGGCCATACAGAACTATGCTATGCTCAGACGAATGTGGA 745  
Qy AGACGGAGAGCCCGAGATTCATGCCACCTTTGTCGGCCCTATGTGGGCAATTCATCTGC 649  
Db AGACGGAGAGCCCGAGATTCATGCCACCTTTGTCGGCCCTATGTGGGCAATTCATCTGC 805  
Qy GCATTGGCTCTGACTGGAAGACGCTTGCCATCTGCTGAGGACGAGCTGCAGGCTGC 709  
Db GCATTGGCTCTGACTGGAAGACGCTTGCCATCTGCTGAGGACGAGCTGCAGGCTGC 865  
Qy ACTTCATGCTCTCTCCAGTGTGTGGGTACAGCGGAGCAGCGGCCCTCTCAGGA 769  
Db ACTTCATGCTCTCTCCAGTGTGTGGGTACAGCGGAGCAGCGGCCCTCTCAGGA 925  
Qy TGACTATGTGCTCTGACCTGAAGGAGATCCAGAGGCTGTGAAGCTCTCTGGGTGAGGT 829  
Db TGACTATGTGCTCTGACCTGAAGGAGATCCAGAGGCTGTGAAGCTCTCTGGGTGAGGT 985  
Qy CGCTGATGCCAGTCTGCTTCTGCTTCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCT 889  
Db CGCTGATGCCAGTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCT 1045  
Qy ACAGCTCTTCAAGGAGAGTGAAGTGTGAGCTTGAAGCTTGAAGCTTGAAGCTTGAAG 949  
Db ACAGCTCTTCAAGGAGAGTGAAGTGTGAGCTTGAAGCTTGAAGCTTGAAGCTTGAAG 1005  
Qy ACTCTATCATCTCTCGGCAAGCGACCACTTTATTTGGCACTGTGCTCTCTCTTCACTG 1009  
Db ACTCTATCATCTCTCGGCAAGCGACCACTTTATTTGGCACTGTGCTCTCTCTTCACTG 1165  
Qy CTTTGTGAAGCGGAGCGGACCTCCAGGAGAGCCGCTCTTCTTCTTCTTCTTCTTCTTCT 1069  
Db CTTTGTGAAGCGGAGCGGACCTCCAGGAGAGCCGCTCTTCTTCTTCTTCTTCTTCTTCT 1225  
Qy GGGCCCTTAAGCTGGGAGAGTCTGATT 1100  
Db GGGCCCTTAAGCTGGGAGAGTCTGATT 1256

RESULT 6  
AAK52494  
ID AAK52494 standard; cDNA; 4850 BP.  
XX

AAK52494;  
06-NOV-2001 (first entry)  
Human polynucleotide SEQ ID NO 2023.  
Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
tissue growth factor; immunomodulatory; cancer; leukaemia;  
nervous system disorder; arthritis; inflammation; ss.  
Homo sapiens.  
XN WO200157190-A2.  
PD 09-AUG-2001.  
PF 05-FEB-2001; 2001WO-US004098.  
XX 03-FEB-2000; 2000US-00496914.  
PR 27-APR-2000; 2000US-00560875.  
PR 20-JUN-2000; 2000US-00598075.  
PR 19-JUL-2000; 2000US-00620325.  
PR 01-SEP-2000; 2000US-00654936.  
PR 15-SEP-2000; 2000US-00663561.  
PR 20-OCT-2000; 2000US-00693325.  
PR 30-NOV-2000; 2000US-00728422.  
XX (HYSE-) HYSEQ INC.  
Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;  
Ma Y, Zhao Q, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
Xue AJ, Yang Y, Wejhrman T, Goodrich R;  
WPI; 2001-476283/51.  
P-PSDB; AAM79361.  
Nucleic acids encoding polypeptides with cytokine-like activities, useful  
in diagnosis and gene therapy.  
Claim 1; Page 4424-4425; 6221pp; English.  
The invention relates to polynucleotides (AAK51456-AAK53435) and the  
encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to  
cytokine, cell proliferation or cell differentiation or which may induce  
production of other cytokines in other cell populations. The  
polynucleotides and polypeptides are useful in gene therapy, vaccines or  
peptide therapy. The polypeptides have various cytokine-like activities,  
e.g. stem cell growth factor activity, haematopoiesis regulating  
activity, tissue growth factor activity, immunomodulatory activity and  
activin/inhibin activity and may be useful in the diagnosis and/or  
treatment of cancer, leukaemia, nervous system disorders, arthritis and  
inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111  
(AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the  
sequence listing were missing at the time of publication  
Sequence 4850 BP; 1142 A; 1218 C; 1244 G; 1246 T; 0 U; 0 Other;  
Query Match 94.7%; Score 1041.4; DB 4; Length 4850;  
Best Local Similarity 99.4%; Pred. No. 2.5e-289;  
Matches 1045; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
Qy 50 GCATGGGGCGCTTTGGGAACAGCGCGATCACTTCTGGGCTCTCTGGCATTTGCAAGC 109  
Db 206 GCCAGGGGCGCTTTGGGAACAGCGCGATCACTTCTGGGCTCTCTGGCATTTGCAAGC 265  
Qy 110 TGTAAACCGTACTTGGCTGTCTCTTGGATTGATACAGCATCAAGCTCTCTT 169  
Db 266 TGCTGAACCGTCTTGGCTGTCTCTTGGATTGATACAGCATCAAGCTCTCTT 325  
Qy 170 TCACCAACCTCCATGTGTCTTACCAAGAGTACTTCAAGCTGGAGCCCTTCCAGGTTACC 229  
Db 326 TCACCAACCTCCATGTGTCTTACCAAGAGTACTTCAAGCTGGAGCCCTTCCAGGTTACC 385

QY 230 ATCCGGTTCATCAGCTTGAGGATTTTCATGGAAGCTGSCACCCACCCACCTGCGCCCTG 289  
 Db |||||  
 QY 386 ATCCGGTTCATCAGCTTGAGGATTTTCATGGAAGCTGSCACCCACCCACCTGCGCCCTG 445  
 Db |||||  
 QY 290 AGAAGCGGTGGCATACTGCTTTGAGGTGGCAGCCAGGAGCCAGATAGAAGACGT 349  
 Db |||||  
 QY 446 AGAAGCGGTGGCATACTGCTTTGAGGTGGCAGCCAGGAGCCAGATAGAAGACGT 505  
 QY 350 GCCCATCAAGGAAGAACCCCTTTGGCCATCTCGGATCAGTTTCATGTAGTTTCA 409  
 Db |||||  
 QY 506 GCCCATCAAGGAAGAACCCCTTTGGCCATCTCGGATCAGTTTCATGTAGTTTCA 565  
 QY 410 ACAAGTCCGAGCTTTTTCAGGCAATTCCTTCAGTGCTTCTTACAGAGAACAAATGGAGCC 469  
 Db |||||  
 QY 566 ACAAGTCCGAGCTTTTTCAGGCAATTCCTTCAGTGCTTCTTACAGAGAACAAATGGAGCC 625  
 QY 470 AGAGATTTTCTCAAAGGAACATCCGCTGCTTGGCCCTGCGAGGAGCCAGCCAGTTCC 529  
 Db |||||  
 QY 626 AGAGATTTTCTCAAAGGAACATCCGCTGCTTGGCCCTGCGAGGAGCCAGCCAGTTCC 685  
 QY 530 CGTCTTAGAAGAACACAGGCACTTACAGAGTACATGATATGTCAGACGAATATGGTGA 589  
 Db |||||  
 QY 686 CGTCTTAGAAGAACACAGGCACTTACAGAGTACATGATATGTCAGACGAATATGGTGA 745  
 QY 590 AGAGCGGAGAGGCCAGATTCATGCCACCTTGTCCGGCCCTATGTGGGCAATTCATCTGC 649  
 Db |||||  
 QY 746 AGAGCGGAGAGGCCAGATTCATGCCACCTTGTCCGGCCCTATGTGGGCAATTCATCTGC 805  
 QY 650 GCATTGGCTCTGACTGGAAGAACGCTGTGCCATGCTGAAGAGCGGACTGAGGCTCGC 709  
 Db |||||  
 QY 806 GCATTGGCTCTGACTGGAAGAACGCTGTGCCATGCTGAAGAGCGGACTGAGGCTCGC 865  
 QY 710 ACTTCATGCTCTCCGAGTGTGTGGCTACAGCCGAGCAGCGGCCCTCTCAGCA 769  
 Db |||||  
 QY 866 ACTTCATGCTCTCCGAGTGTGTGGCTACAGCCGAGCAGCAGCGGCCCTCTCAGCA 925  
 QY 770 TGACTATGCTGCTGCTCAGCTGAAGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGGT 829  
 Db |||||  
 QY 926 TGACTATGCTGCTGCTCAGCTGAAGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGGT 985  
 QY 830 CGCTGGATGCCAGTGGTCTACGTTGCTTACTGATTCGAGAGTTATGTGCTGAGCTCC 889  
 Db |||||  
 QY 986 CGCTGGATGCCAGTGGTCTACGTTGCTTACTGATTCGAGAGTTATGTGCTGAGCTCC 1045  
 QY 890 AACAGCTCTTCAAAGGGAAGGTGAAGTGTGTGAGCTGAAGCTGAGGTGCCCAGGTCG 949  
 Db |||||  
 QY 1046 AACAGCTCTTCAAAGGGAAGGTGAAGTGTGTGAGCTGAAGCTGAGGTGCCCAGGTCG 1105  
 QY 950 ACCTGTACATCTCGGCCAAGCCGACACTTATTTGGCACTGTGCTCTCTCTCACTG 1009  
 Db |||||  
 QY 1106 ACCTGTACATCTCGGCCAAGCCGACACTTATTTGGCACTGTGCTCTCTCTCACTG 1165  
 QY 1010 CTTTGTGAAGCGGAGCGGAGCACTTCCAGGGAGGCGCTCTTCTTCTTCTGGCATGGACA 1069  
 Db |||||  
 QY 1166 CTTTGTGAAGCGGAGCGGAGCACTTCCAGGGAGGCGCTCTTCTTCTTCTGGCATGGACA 1225  
 QY 1070 GGCCCCCTTAAGCTGCGGAGCAGTTCGATT 1100  
 Db |||||  
 QY 1226 GGCCCCCTTAAGCTGCGGAGCAGTTCGATT 1256

## RESULT 7

AAV65634

ID AAV65634 standard; cDNA; 5009 BP.

XX AC AAV65634;

XX AC AAV65634;

DT 16-DEC-1998 (first entry)

 XX First EcoRI nucleotide fragment of human KIAA0180.  
 XX O-fucosyltransferase; epidermal growth factor; EGF; glycosylation;

KW O-fucose; inhibitor; sensory neuron; retinal neuron; human; heart; ss.  
 OS Homo sapiens.  
 XX WO9833924-A1.  
 XX PD 06-AUG-1998.  
 XX PF 17-DEC-1997; 97WO-US023401.  
 XX PR 31-JAN-1997; 97US-00792498.  
 XX PR 26-NOV-1997; 97US-00978741.  
 XX PA (GETH ) GENENTECH INC.  
 XX PI Wang Y, Spellman MW;  
 XX WI PI; 1998-437477/37.  
 XX DR P-PSDB; AAW80577.

Human O-fucosyltransferase able to glycosylate epidermal growth factor domains - useful for diagnosis and treatment of diseases involving overexpression of the enzyme.

Example 1; Fig 11; 90pp; English.

This represents a first EcoRI nucleotide fragment of human KIAA0180. This 5009 basepair partial cDNA encodes for a protein of unknown function from myelast cell line KG-1. The invention provides a human heart O-fucosyltransferase that can glycosylate an epidermal growth factor (EGF) domain of a polypeptide with an activated O-fucose residue. Inhibitors of O-fucosyltransferase, e.g. mutants with increased affinity for the EGF domains, are used in diagnosis and treatment of conditions associated with overexpression of O-fucosyltransferase, to promote survival of sensory (retinal) neurons. Probes based on EGF domain polypeptide are used to detect gene amplification and expression. The expression can also be determined at the protein level using antibodies specific for O-fucosyltransferase

Sequence 5009 BP; 1235 A; 1195 C; 1268 G; 1311 T; 0 U; 0 Other;

Query Match 93.9%; Score 1033.4; DB 2; Length 5009;  
 Best Local Similarity 99.9%; Pred. No. 5.1e-287;  
 Matches 1034; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 66 GAACAGCGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAGCTGTCTAAACCGTACCTT 125  
 Db 1 GAACAGCGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAGCTGTCTAAACCGTACCTT 60  
 QY 126 GGCTGTCCCTCTTGGATTTGATGATCAGCATCAAGGCTCTTTCACCACTCCATGT 185  
 Db 61 GGCTGTCCCTCTTGGATTTGATGATCAGCATCAAGGCTCTTTCACCACTCCATGT 120  
 QY 186 GTCTTACAGAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATGGGTTCATCAGCTT 245  
 Db 121 GTCTTACAGAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATGGGTTCATCAGCTT 180  
 QY 246 GGAGGATTTTCATGGAAGCTGGCACCACCCACTGGCCCTCGAGAGCGGGTGGCATA 305  
 Db 181 GGAGGATTTTCATGGAAGCTGGCACCACCCACTGGCCCTCGAGAGCGGGTGGCATA 240  
 QY 306 CTGCTTTGAGTGGCAGCCGAGCCGAGCAAGAGAGCTGCCCTCCATGAAGGAAGG 365  
 Db 241 CTGCTTTGAGTGGCAGCCGAGCCGAGCAAGAGAGCTGCCCTCCATGAAGGAAGG 300  
 QY 366 AAACCCCTTTGGCCATTTCTGGGATCAGTTTCATGTAGTTTCAACAGTCGGAGCTTTT 425  
 Db 301 AAACCCCTTTGGCCATTTCTGGGATCAGTTTCATGTAGTTTCAACAGTCGGAGCTTTT 360  
 QY 426 TACAGGCAATTTCTTCAAGTGTCTTCTTACAGAGAAATGGAGCCAGAGATTTTCTCCAAA 485  
 Db 361 TACAGGCAATTTCTTCAAGTGTCTTCTTACAGAGAAATGGAGCCAGAGATTTTCTCCAAA 420



Db 16 CACTTTGGCCGAAACCTACCGGGCATTCATCTGCGCATTTGGCTCTGACTGGAAGAACGCC 75  
Qy 676 TGTGTCATCTGAAGGACGGAGCTGACGCTCGCACTTCATGGCTCTCCGCACTGCTG 735  
Db 76 TGTGTCATCTGAAGGACGGAGCTGACGCTCGCACTTCATGGCTCTCCGCACTGCTG 135  
Qy 736 GGTACAGCCGACGACAGCGGCGCCCTCAGCATGACTATGTGCTGCTGCTGACTGAAG 795  
Db 136 GGTACAGCCGACGACAGCGGCGCCCTCAGCATGACTATGTGCTGCTGACTGAAG 195  
Qy 796 GAGATCCAGAGGCTGTGAAGCTCTGGGTGAGTCTGCTGAGTGCCTGCTGCTGCTGCTG 855  
Db 196 GAGATCCAGAGGCTGTGAAGCTCTGGGTGAGTCTGCTGAGTGCCTGCTGCTGCTGCTG 255  
Qy 856 GCTACTGATTCGAGAGTATGTGCTGCTGAGCTCAGAGCTCTTCAAGAGGAAGTGAAG 915  
Db 256 GCTACTGATTCGAGAGTATGTGCTGAGCTCAGAGCTCTTCAAGAGGAAGTGAAG 315  
Qy 916 GTGTGAGCCTGAAGCCTGAGGTGCGCCAGGTGCACTGTATCTCTCGGCCAAGCCGAC 975  
Db 316 GTGTGAGCCTGAAGCCTGAGGTGCGCCAGGTGCACTGTATCTCTCGGCCAAGCCGAC 375  
Qy 976 CACTTTATGGAACCTGTCT 1035  
Db 376 CACTTTATGGAACCTGTCT 435  
Qy 1036 CAGGGAGGCGCT 1095  
Db 436 CAGGGAGGCGCT 495  
Qy 1096 TGATT 1100  
Db 496 TGATT 500

RESULT 9  
ADQ56824  
ID ADQ56824 standard; DNA; 610 BP.  
XX ADQ56824;  
AC ADQ56824;  
XX 21-OCT-2004 (first entry)  
XX Novel canine microarray-related DNA sequence SeqID8126.  
XX canine microarray; drug screening; toxicity assay;  
KW environmental pollutant; cellular response; gene expression profile;  
KW toxic response; liver necrosis; fatty liver disease;  
KW protein adduct formation; hepatitis; dog; ds.  
XX Canis familiaris.  
XX WO2004063324-A2.  
XX 29-JUL-2004.  
XX 05-MAY-2003; 2003WO-US013853.  
XX 03-MAY-2002; 2002US-0377240P.  
XX (GENE-) GENE LOGIC INC.  
XX (PFIZ ) PFIZER PROD INC.  
XX Diggins JC, Porter M, Wei T;  
XX WPI; 2004-561890/54.  
XX New isolated nucleic acid molecule, useful for drug screening and  
PT toxicity assays or for assessing the impact, including toxicity, of a  
PT compound, pharmaceutical agent or environmental pollutant on a cell or  
PT living organism.  
XX

PS Claim 1; SEQ ID NO 8126; 41pp; English.  
XX This invention is related to a novel isolated canine nucleic acid  
CC sequences and the construction of canine microarrays containing a  
CC significant portion of the canine genome. The isolated canine nucleic  
CC acid sequences of the invention may be useful for drug screening and  
CC toxicity assays. The invention is therefore useful for assessing the  
CC impact, including toxicity, of a compound, pharmaceutical agent or  
CC environmental pollutant on a cell or living organism. The methods are  
CC useful for detecting genes that are up- or down-regulated in canines in a  
CC disease state. The sequences are useful as diagnostic agents or markers  
CC to detect a cellular response in a sample individually or as part of a  
CC gene expression profile. It is also useful as a target for agents that  
CC modulate gene expression or activity. The database is useful for  
CC producing electronic Northern blots that allow the user to determine the cell  
CC type or tissue in which a given gene is expressed and to allow  
CC determination of the abundance or expression level of a given gene in a  
CC particular tissue or cell. The methods are useful for determining the  
CC similarity of a toxic response to one or more individual compounds. The  
CC methods are useful for predicting at least one toxic response or the  
CC likelihood that a compound or test agent will induce various specific  
CC pathologies such as those of the liver (liver necrosis, fatty liver  
CC disease, protein adduct formation or hepatitis), those of the kidney,  
CC heart, brain or testes, or other pathologies associated with at least one  
CC of the toxins. The methods are also useful for predicting or elucidating  
CC the potential cellular pathways influenced, induced or modulated by the  
CC compound or test agent due to the similarity of the expression profile  
CC compared to the profile induced by a known toxin. The present sequence is  
CC that of a canine DNA sequence which was claimed for use during the  
CC production of a canine microarray of the invention.  
XX SQ Sequence 610 BP; 155 A; 171 C; 149 G; 129 T; 0 U; 6 Other;

Query Match 36.9%; Score 406.2; DB 13; Length 610;  
Best Local Similarity 89.0%; Pred. No. 1.8e-106;  
Matches 446; Conservative 0; Mismatches 54; Indels 1; Gaps 1;  
Qy 171 CACCAACTCCTCCTGCTCTTACCAGAGAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCA 230  
Db 1 CACCAATCTCCTGCTCTTACCAGAGAGTACTTCAAGCTNNAGCCCTCAGGNTACCA 60  
Qy 231 TCGGCTCATCAGCTTGGAGGATTTTCATGGAGAGCTGGCACCCTGCTGCCCCCTGA 290  
Db 61 TCGGCTTATCAGCTTGGAGGATTTTCATGGAGAGCTGGCACCCTGCTGCCCCCTGA 120  
Qy 291 GAAGCGGTGGCATACTGCTTTGAGTGGAGCCAGCCAGGAGCCAGATAGAGACGTG 350  
Db 121 GAGCGAGTGGCATACTGCTTTGAAGTGGAGCCAGCCAGGAGCCCTGATAGAGACATG 180  
Qy 351 CCCCATGAAGGAAGAAACCCCTTTGGCCCATCTTGGGATCAGTTTCATGTGAGTTTCAA 410  
Db 181 CCCCATGAAGGAAGAAATCCCTTTGGCCCATTTTGGGATCAGTTTCATGTGAGTTTCAA 240  
Qy 411 CAAGTCGGAGCTTTTTCAGGCAATTTCTTCAGTGTCTTCTACAGAGAACATGAGCCA 470  
Db 241 CAAGTCCAAAGCTTTTTCAGGCAATTTCTTCAGTGTCTTCTACAAAGACCAATGGATTCA 300  
Qy 471 GAGATTTTCTCAAAGGAACATCCAGTGTCTTGGCCCTGCGAGGAGCCCGCCAGTTCCC 530  
Db 301 GAGATTTTCTCAAAGGAACATCCAGTGTCTTGGCCCTGCGAGGAGCCCGCCAGTTCCC 359  
Qy 531 CGTCTAGAGAACACAGGCGCCTACAGAGTACATGATGTGTCAGAGAAATGGTGAA 590  
Db 360 CGTCTAGAGAACACAGGCGCCTACAGAGTACATGATGTGTCAGAGAAATGGTGAG 419  
Qy 591 GACGGAGAGGCGCCAGATTTTCATGCCACCTTGTCCGGCCCTATGTGGGCAATTCATCTCG 650  
Db 420 GACGGAGAGGCGCCAGATTTTCATGCCACCTTGTCCGGCCCTATGTGGGCAATTCATCTCG 479  
Qy 651 CATTGGCTCTGACTGGAGAA 671  
Db 480 CATTGGCTCTGACTGGATGGA 500

RESULT 10  
AAK79915  
ID AAK79915 standard; DNA; 10331 BP.  
XX  
AC AAK79915;  
XX  
DT 07-NOV-2001 (first entry)  
XX  
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:34727.  
XX  
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
KW cytostatic; gene therapy; vaccine; metastasis; ds.  
XX  
XX Homo sapiens.  
XX  
XX WO200157182-A2.  
XX  
XX 09-AUG-2001.  
XX  
XX 17-JAN-2001; 2001WO-US001354.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205115P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 11-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0225119P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226688P.  
PR 23-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229309P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 09-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
XX 21-SEP-2000; 2000US-0234274P.  
XX 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 28-SEP-2000; 2000US-0236327P.  
PR 28-SEP-2000; 2000US-0236367P.  
PR 28-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.



PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234224P.  
PR 25-SEP-2000; 2000US-0234377P.  
PR 25-SEP-2000; 2000US-0234378P.  
PR 25-SEP-2000; 2000US-0234384P.  
PR 25-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 01-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 01-DEC-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.

PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-465570/50.  
XX Isolated nucleic acid molecule encoding a reproductive system antigen is  
XX used in preventing, treating or ameliorating a medical condition.  
XX Disclosure; SEQ ID NO 6032; 1297pp + Sequence Listing; English.  
XX The present invention provides the protein and coding sequences of a  
XX number of human reproductive system related antigens. These can be used  
XX in the prevention and treatment of reproductive system disorders, including a  
XX including cancer. The present sequence is a genomic sequence encoding a  
XX protein of the invention  
XX SQ Sequence 10331 BP; 2718 A; 2405 C; 2599 G; 2609 T; 0 U; 0 Other;  
Query Match 22.8%; Score 251.2; DB 4; Length 10331;  
Best Local Similarity 97.0%; Pred. No. 3.4e-61;  
Matches 256; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
QY 656 GCTCTGACTGGAAGAAGCGCTGTGCCATGCTGAAGGACGGGACTGACAGCTCGCACTTCA 715  
Db |||||  
QY 716 TGGCTCTCCGAGTGTGGCTACAGCCGAGCAGCAGAGGGCCCCCTCAGCATGACTA 775  
Db |||||  
QY 776 TGTGCTGCTGACTGAGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGTGGCTGG 835  
Db |||||  
QY 836 ATGCCAGTTCGCTTACGTTGCTACTGATCCGAGATGATGCTGAGCTCCACACAGC 895  
Db |||||  
QY 896 TCTTCAAAGGGAAGGTGAAGTGG 919  
Db |||||  
ID ACH32225 standard; cDNA; 479 BP.  
XX ACH32225;  
XX  
XX 13-OCT-2003 (first entry)  
XX Human endothelial cell cDNA #358.  
XX Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;  
XX genome mapping; biodiversity; genetic disorder.  
XX Homo sapiens.  
XX US2003073623-A1.  
XX 17-APR-2003.  
XX 30-JUL-2001; 2001US-00918995.  
XX



```

PR 30-JUL-2001; 2001US-00918995.
XX (DRMA/) DRMANAC R T.
PA (LABA/) LABAT I.
PA (STAC/) STACHE-CRAIN B.
PA (DICK/) DICKSON M C.
PA (JONE/) JONES L W.
XX
XX Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;
XX WPI; 2003-615964/58.
XX
XX New polynucleotide sequences obtained from various cDNA libraries, useful
PT as hybridization probes, as oligomers for PCR, for chromosome and gene
PT mapping, in the recombinant production of protein, or in generating
PT antisense DNA or RNA.
XX
XX Claim 1; SEQ ID NO 19437; 44pp; English.
XX
XX The invention relates to an isolated polynucleotide comprising any one of
CC 38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was
CC determined by the technique of SBH (sequencing by hybridisation). Also
CC included is a purified polypeptide comprising a sequence corresponding to
CC a reading frame of the novel polynucleotide. The nucleic acid sequences
CC are useful in diagnostics as expressed sequence tags (EST) for
CC identifying expressed genes or for physical mapping of the human genome,
CC in forensics, in assessing biodiversity, or in identifying mutations
CC responsible for genetic disorders and other traits. The nucleotide
CC sequences are also useful as hybridisation probes, as oligomers for PCR,
CC for chromosome and gene mapping, in the recombinant production of
CC protein, or in generating antisense DNA or RNA. The purified polypeptide
CC is useful for generating antibodies specific for it. The present sequence
CC is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?DocID=20030073623
XX
XX Sequence 479 BP; 99 A; 140 C; 143 G; 97 T; 0 U; 0 Other;
XX
XX Query Match 21.2%; Score 233; DB 9; Length 479;
XX Best Local Similarity 100.0%; Pred. No. 1.5e-56;
XX Matches 233; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 868 GAGAGTTATGTCCTGAGCTCCACAGCTTTCAAAGGGAAGGTGAGTGGTGGCTG 927
XX 31 GAGAGTTATGTCCTGAGCTCCACAGCTTTCAAAGGGAAGGTGAGTGGTGGCTG 90
XX
XX 928 AAGCCTGAGTGGCCAGGTGCACTGTACATCTCTCGCCCAAGCCGACCACTTTATTGGC 987
XX 91 AAGCCTGAGTGGCCAGGTGCACTGTACATCTCTCGCCCAAGCCGACCACTTTATTGGC 150
XX
XX 988 AACTGTGTCTCTCTTCACTGCCTTTGTGAGCGGGAGCGGACCTCCAGGGGAGGCGG 1047
XX 151 AACTGTGTCTCTCTTCACTGCCTTTGTGAGCGGGAGCGGACCTCCAGGGGAGGCGG 210
XX
XX 1048 TCTTCTTTCTTCGGCATGCAAGCGCCCTTAAGCTGCGGACGAGTTCCTGATT 1100
XX 211 TCTTCTTTCTTCGGCATGCAAGCGCCCTTAAGCTGCGGACGAGTTCCTGATT 263
XX
XX RESULT 13
XX ACH35532
XX ID ACH35532 standard; cDNA; 477 BP.
XX
XX ACH35532;
XX
XX 13-OCT-2003 (first entry)
XX
XX Human endothelial cell cDNA #3665.
XX
XX Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;
XX genome mapping; biodiversity; genetic disorder.
XX

```

```

OS Homo sapiens.
XX US2003073623-A1.
XX
XX 17-APR-2003.
XX
XX 30-JUL-2001; 2001US-00918995.
XX
XX 30-JUL-2001; 2001US-00918995.
XX
XX (DRMA/) DRMANAC R T.
XX (LABA/) LABAT I.
XX (STAC/) STACHE-CRAIN B.
XX (DICK/) DICKSON M C.
XX (JONE/) JONES L W.
XX
XX Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;
XX WPI; 2003-615964/58.
XX
XX New polynucleotide sequences obtained from various cDNA libraries, useful
PT as hybridization probes, as oligomers for PCR, for chromosome and gene
PT mapping, in the recombinant production of protein, or in generating
PT antisense DNA or RNA.
XX
XX Claim 1; SEQ ID NO 22744; 44pp; English.
XX
XX The invention relates to an isolated polynucleotide comprising any one of
CC 38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was
CC determined by the technique of SBH (sequencing by hybridisation). Also
CC included is a purified polypeptide comprising a sequence corresponding to
CC a reading frame of the novel polynucleotide. The nucleic acid sequences
CC are useful in diagnostics as expressed sequence tags (EST) for
CC identifying expressed genes or for physical mapping of the human genome,
CC in forensics, in assessing biodiversity, or in identifying mutations
CC responsible for genetic disorders and other traits. The nucleotide
CC sequences are also useful as hybridisation probes, as oligomers for PCR,
CC for chromosome and gene mapping, in the recombinant production of
CC protein, or in generating antisense DNA or RNA. The purified polypeptide
CC is useful for generating antibodies specific for it. The present sequence
CC is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?DocID=20030073623
XX
XX Sequence 477 BP; 95 A; 141 C; 142 G; 97 T; 0 U; 2 Other;
XX
XX Query Match 20.2%; Score 221.8; DB 9; Length 477;
XX Best Local Similarity 96.3%; Pred. No. 2.5e-53;
XX Matches 237; Conservative 0; Mismatches 8; Indels 1; Gaps 1;
XX
XX 855 TGCTACTGATTCCGAGAGTTATGTGCTGAGCTCCACAGCTCTTCAAAGGAAGGTGAA 914
XX 11 TGATGCTGAGCNCCTAGAGATATGTGCTGAGCTCC-ACAGTTCTTCAAAGGAAGGTGAA 69
XX
XX 915 GGTGCTGAGCCTGAAGCTGAGGTGGCCAGCTGACCTGACATCTCTCGCCCAAGCCGA 974
XX 70 GGTGCTGAGCCTGAAGCTGAGGTGGCCAGCTGACCTGACATCTCTCGCCCAAGCCGA 129
XX
XX 975 CCACCTTTATTGGCAACTGTGTCTCTCTTCACTGCTCTTGTGAAGCGGAGCGGACCT 1034
XX 130 CCACCTTTATTGGCAACTGTGTCTCTCTTCACTGCTCTTGTGAAGCGGAGCGGACCT 189
XX
XX 1035 CCAGGGAGGCGGCTCTTCTTTCTTCGGCATGGAAGGCCCCCTTAAGCTGCGGGACGATT 1094
XX 190 CCAGGGAGGCGGCTCTTCTTTCTTCGGCATGGAAGGCCCCCTTAAGCTGCGGGACGATT 249
XX
XX 1095 CTGATT 1100
XX 250 CTGATT 255
XX
XX RESULT 14

```

ABLI13947  
ID ABL13947 standard; cDNA; 1209 BP.  
AC ABL13947;  
XX  
XX 26-MAR-2002 (first entry)  
XX  
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 36323.  
XX  
XX Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ss.  
XX  
XX Drosophila melanogaster.  
XX  
XX WO200171042-A2.  
XX  
XX PD 27-SEP-2001.  
XX  
XX PF 23-MAR-2001; 2001WO-US009231.  
XX  
XX PR 23-MAR-2000; 2000US-0191637P.  
XX  
XX PR 11-JUL-2000; 2000US-00614150.  
XX  
XX PA (PEKE ) PE CORP NY.  
XX  
XX PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
XX DR WPI: 2001-656860/75.  
XX  
XX DR P-PSDB; AB569844.  
XX  
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more  
XX genes from Drosophila and for elucidating cell signaling and cell-cell  
XX interactions.  
XX  
XX PS Claim 1; SEQ ID NO 36323; 21pp + Sequence Listing; English.  
XX  
XX CC The invention relates to an isolated nucleic acid detection reagent  
XX capable of detecting 1000 or more genes from drosophila. The invention is  
XX useful in developmental biology and in elucidating cell signalling and  
XX cell-cell interactions in higher eukaryotes for the development of  
XX insecticides, therapeutics and pharmaceuticals. The invention  
XX discloses genomic DNA sequences (ABLI16176-ABLI30511), expressed DNA  
XX sequences (ABLI01840-ABLI16175) and the encoded proteins (ABH5773-  
XX ABH72072). The sequence data for this patent did not form part of the  
XX printed specification, but was obtained in electronic format directly  
XX from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX SQ Sequence 1209 BP; 302 A; 313 C; 307 G; 287 T; 0 U; 0 Other;  
XX  
XX Query Match 15.1%; Score 166.4; DB 4; Length 1209;  
XX Best Local Similarity 51.5%; Pred. No. 3.7e-37;  
XX Matches 563; Conservative 0; Mismatches 486; Indels 45; Gaps 6;  
XX  
XX QY 17 GGGACCCGGCGTTACTGCTCTACTGCTCCCTGATGGGCGCTTTGGGACCGAGCGG 76  
XX DB 80 GCGATCCCAATGGCTACTCACTACTGTCTCGTGTATGGAGCGCTTGGACCGAGCGG 139  
XX  
XX QY 77 ATCACTTCTGGGCTCTGGCAATTTGCAAGCTCTAAACCTGACCTGGCTGCTCCCTC 136  
XX DB 140 ACCACTTCTGGGATCATTTGGCTTCCCAAGCGCTTATCGACCTGATCTGCTGCGC 199  
XX  
XX QY 137 CTPTGGATTGAGTACGAGCATCAACAGCTCTTTTCCCAACCTCCATGTGCTCTACAGA 196  
XX DB 200 CGTGGGTGGAGT-----ATCGTAGGGGTGAATCGATCCCGCTCAGGTACCGTTCA 253  
XX  
XX QY 197 AGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGTCATCAGCTTGGAGGATTTCA 256  
XX DB 254 CATATTTTGAAGTGAGGCCCCGTAAGGAATACCATCGGTCATCACCATGGCAGATTTCA 313  
XX  
XX QY 257 TGGAGAAAGCTGGACCCACCTGCGCCCTCGAGAAGCGGGTGGCATACTGCTTTGAGG 316  
XX DB 314 TGTGGCACCTGGCCGACACATTTGGCCGAATTCGAGCGAGTGTCTATTTTGTCTACA 373

QY 317 -----TGGCAGCCCGAGCGAGCCAGATAGAAGAGCGTGCCCCCATGAAGG 361  
DB 374 AACGATATAGCTTTACAGCAGGAGAGAGACGATCCAGACAAGCCCAATTTGCCAGCGCAAGG 433  
QY 362 AAGGAAACCCCTTTGGCCCAATTTCTGGGATCAGTTTTCATGTGAGTTTCAACAAGTCGGAGC 421  
DB 434 ATGGCAATCTCTTTTGGTCCCTTTTGGGACACTTTTTCACATAGACTTTGTGCGGTGAGAGT 493  
QY 422 TTTTTCACAGGCAATTTCTTTCAGTGTCTTCTACAG-----AGAACATGAGGCCAGA 472  
DB 494 TCTATGCCCACTTTCAATTTTGAATGTGCATCATAGCAACAGAGGCTGCCAAGTGGCAGACCA 553  
QY 473 GATTTTCTCCAAAGGAACATCCGGTGTCTGCTTGCCTGCGAGGAGCCCGAGCCAGTTTCCCGC 532  
DB 554 AATATCTGTGAGATCATATCCGTAATCCGTAATCCGTAATCCGTAATCCGTAATCCGTAAT 613  
QY 533 TCCTAGAAGAACACAGGCCACTACAGAAAGTACATGATGTGTGAGTGTGAGTGTGAGTGTGAG 592  
DB 614 TTACGTAGAACTGTCAAGCTGTGAGCGTACTTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 673  
QY 593 CGGAGAGGCCAGATTCATGCCACCTTGTCCCGC-----CCTATGTGGGATTCATCTGC 649  
DB 674 CATCTAAGGATTTTATCCGAGAGAGTTCGCTCGGGGTGCTTTTGGGCAATTCATCTGC 733  
QY 650 GCATTGGCTCTGACTGGAAGAACGCTGTGCCATGTGTGAGGAGCG-----GACTG 700  
DB 734 GCACGGTATCGATTGGGTGAGAGCGCTGTGAGCAGCTCAAGGATAGCCAGCATCTGTGTTG 793  
QY 701 CAGGCTGCGCACTTTCATGGCCTCTCCGAGTGTGTGGGCTACAGCCGCGAGCAGCGGCC 760  
DB 794 CTTGCGCGCAGTGTCTGGGCTATAAATAAGAGTGTGTGCACTTACCCGGAGCTCTGCA 853  
QY 761 CCTCAGCATGACTATGTGCTGCTGACCTGAAGAGATCCAGAGGGCTGTGAAGCTCT 820  
DB 854 TGCCCTCCAAAGGAGCGATCATCCGACCTAAAGAGAACCATTAAGAACGTGCGCCAA 913  
QY 821 GGGTGAAGTCCGCTGATGCGCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 880  
DB 914 CTCAGCCGAGCAACGAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCA 973  
QY 881 CTGAGCTCCACAGCTCTTCAAGGAGAGTGAAGTGTGAGGCTGAGGCTGAGGCTGAGGCTG 940  
DB 974 GTGACTTAAACACGCGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1033  
QY 941 ---CCAGGTGACCTGTATCATCTCGGCAAGCGGACACTTTATTTGGCACTGTGCT 997  
DB 1034 ATCTTTACCTGAGCTTGGCAATTCGAGACGTGGAACCACTTTATCGCACTGTATAT 1093  
QY 998 CCTCTTCACTGCTCTTTTGAAGCGGAGCGGACCTCCAGGAGGCGCTCTTCTTCT 1057  
DB 1094 CTTCTTACTCGGCATTCGAAAAAAGGGAACGAGATGTGACGCTTTTCCATCGTACTTCT 1153  
QY 1058 TCGGCAATGGACAGG 1071  
DB 1154 GGGGATTCGCCCAAG 1167  
XX  
XX RESULT 15  
XX ID ABL13946/C  
XX ID ABL13946 standard; cDNA; 3264 BP.  
XX AC ABL13946;  
XX XX  
XX DT 26-MAR-2002 (first entry)  
XX  
XX DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 36320.  
XX  
XX KW Drosophila; developmental biology; cell signalling; insecticide;  
XX KW pharmaceutical; gene; ss.  
XX  
XX OS Drosophila melanogaster.  
XX  
XX PN WO200171042-A2.

XX 27-SEP-2001.  
XX 23-MAR-2001; 2001WO-US009231.  
XX 23-MAR-2000; 2000US-0191637P.  
XX 11-JUL-2000; 2000US-00614150.  
XX (PEXE ) PE CORP NY.  
XX Venter JC, Adams M, Li PWD, Myers EW;  
XX WPI; 2001-656860/75.  
XX P-PSDB; ABB69843.  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
XX genes from Drosophila and for elucidating cell signaling and cell-cell  
XX interactions.  
XX Claim 1; SEQ ID NO 36320; 21pp + Sequence Listing; English.  
XX The invention relates to an isolated nucleic acid detection reagent  
XX capable of detecting 1000 or more genes from Drosophila. The invention is  
XX useful in developmental biology and in elucidating cell signalling and  
XX cell-cell interactions in higher eukaryotes for the development of  
XX insecticides, therapeutics and pharmaceutical drugs. The invention  
XX discloses genomic DNA sequences (AB116176-AB130511), expressed DNA  
XX sequences (AB101840-AB116175) and the encoded proteins (ABB57737-  
XX ABB72072). The sequence data for this patent did not form part of the  
XX printed specification, but was obtained in electronic format directly  
XX from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 3264 BP; 904 A; 742 C; 758 G; 860 T; 0 U; 0 Other;  
XX  
Query Match 13.7%; Score 150.4; DB 4; Length 3264;  
Best Local Similarity 50.9%; Fred. No. 2.4e-32;  
Matches 541; Conservative 0; Mismatches 476; Indels 45; Gaps 6;  
XX 49 TGCATGGGGCGCTTTGGGAACAGGCCGATCACTCTCTGGGCTCTCTGGCATTTGGCAAAG 108  
DB 2098 TTCAAAGGACGCTTTGGCAACAGGCCGACCACTCTCTGGGATCATTTGGCCCTTCGCCAAG 2039  
XX 109 CTGCTAAACCGTACCTTTGGCTGTCCCTCTTGGATTGAGTACCAGCATCAAGCCCTCCT 168  
DB 2038 GCGCTTAAATCGACACCTGATCTCGCCGCGTGGTGGAGT-----ATCGTAGGGTGAA 1985  
XX 169 TTCACCAACCTCCATGTCTCTACAGAACTTTCAGCTGGAGCCCTCCAGCCTTAC 228  
DB 1984 CTGGATCCCGTCAAGGTACCGTTCACACATATTTTGAAGTGGAGCCCTGAAGGAATAC 1925  
XX 229 CATCGGTCATCAGCTTGGAGATTTTCATGAGAACTGGCAACCCACCCACTGGCCCTCCT 288  
DB 1924 CATCGGTCATCACCATGGCAGATTTTCATGTGGCACCTGGCCGAGGACATTTGGCCAGAA 1865  
XX 289 GAGAGCGGGTGGCATACTGCTTTGAGG-----TGGCAGCCCGCAGCAAGC 333  
DB 1864 TCGGAGCGAGTGTCAATTTTGTACAAGGAACGATATAGCCTTCAGCAGGAAGAACGAT 1805  
XX 334 CCAGATAAGAGACGTGCCCCATGAGGAAGAACCCCTTTGGCCCATTCCTGGATCAG 393  
DB 1804 CCAGACAAGCCCAATTGCGACGCCAAGGATGGCAATCCTTTTGGTCCCTTTTGGGCACT 1745  
XX 394 TTTTCATGTGAGTTTCAACAAGTCGGAGCTTTTTCAGGCAATTTCTTCAGTGTCTTCCTAC 453  
DB 1744 TTTTCATAGACTTTTGGCGGTGAGATTTCTATGGCCACTTTTATGATGTGCATCAT 1685  
XX 454 AG-----AGAACATATGGAGCCAGAGATTTTTCCTCAAGGAACATCCGGTGTCTGCC 504  
DB 1684 AGCAACGAGGCTGCCAAGTGGCAGACCAATATCTCTGAGAATCATATCCCGTACTCGCG 1625  
XX 505 CTGCGAGGAGCCCGAGCCAGTTCCTCGCTCTAGAGAACACAGGCCACTACAGAAATAC 564  
DB 1624 TTCACCGGAGCTCGGGTGTAGTTTCTCTGCTTAGAGAACTGCAAGCTGCAGCGTAC 1565

QY 565 ATGGTATGTCAGACAAATGGTGAAGACGGAGAGGCCAGATTCATGCCCCACCTTGTGC 624  
DB 1564 TTGCAGTGGAGTCAACGGTATAGGAAGCATCTAAGGATTTTCATCCGAGAGCAGTTGCC 1505  
QY 625 CGG---CCCTATGTGGGCATTCATCTGCGCATTTGGCTCTGACTGGAAGAACGCTTGTGCC 681  
DB 1504 CGGGTGGCTTTTGGGCATTCATCTGCGCAACGGTATCGATTGGGTGAGAGCCTTGTGAG 1445  
QY 682 ATGCTGAGGACGG-----GACTGCGAGGCTCGCACTTTCATGGCCTCTCCGCAAGTGT 732  
DB 1444 CAGGTCAAGGATAGCCAGCATCTGTTTGCCTCGCGCAGTCTTGGGCTATAAAAAATGAA 1385  
QY 733 GTGGCTACAGCCGAGCAGCAGCGGCCCTCCCTCAGCATGATCATGTGCTGCTGCCTGACCTG 792  
DB 1384 CGTGGTGCACCTCTACCCGGAGCTCTGCATGCGCTTCAAGGAGGCGCATCATCGCCAGCTA 1325  
QY 793 AAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTGCTGGATGCCCAAGTCGGTCTTAC 852  
DB 1324 AAGAGAACCATTAAGAACGTCGCGCAAACTCAGCGCGGACAAACGAAATCAAAATCAGTTTC 1265  
QY 853 GTTCTACTGATTCGGAGAGTTATGTGCTGAGCTCCACAGCTTTCAAAGGGAAGGTG 912  
DB 1264 GTGGCTCAGACTCCAATCAATGATTTGGTGAACATAAACACGCGCCCTTAGTGCATGGGC 1205  
QY 913 AAGGTGGTGAAGCTTGAAGCCCTGAGGTGG---CCAGGTGCGACCTGTATCATCTCCGCCAA 969  
DB 1204 ATCAGTGTGCACAAAGCTGCCGAGGATGATCTTACCTGGACTTGGCCATTTCTCGGACAG 1145  
QY 970 GCCAACCACCTTTATGGCAACTGTGTCTCTCTTCACTGCTTTGTGAAGCGGAGCGG 1029  
DB 1144 TCGAACCACTTTATCGGCAACTGTATATCTCTTACTCGGCATTCGAAAAAAGGGAACGA 1085  
QY 1030 GACCTCCAGGAGGCGCTCTTCTTCTTCTGCGCATGACAGG 1071  
DB 1084 GATGTGACCGGTTTTTCCATCGTACTTCTGGGGATTCGCCAAG 1043

Search completed: November 9, 2005, 04:43:20  
Job time : 644.158 secs

THIS PAGE BLANK (USPTO)

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	1096.8	99.7	1151	3	CR617950	CR617950 full+length
2	892.8	81.2	1141	5	BR353274	BR353274 BX353274
3	885.4	80.5	3611	3	AK081059	AK081059 Mus muscu
4	884.2	80.4	2616	3	AK044629	AK044629 Mus muscu
5	743.6	67.6	1091	5	BM909055	BM909055 AGENCOURT
6	718.4	65.3	841	5	BR345989	BR345989 BX345989
7	717.4	65.2	719	1	AL135434	AL135434 DKF2p7620
8	712.2	64.7	889	4	BI524031	BI524031 603052202
9	706.6	64.2	898	5	EX449372	EX449372 BX449172
10	673.2	61.2	942	7	CO774833	CO774833 ILLUMIGEN
11	662.2	60.2	808	4	BG747760	BG747760 602705255
12	660.2	60.0	801	5	BP144264	BP144264 BP144264
13	639	58.1	988	5	BQ964179	BQ964179 AGENCOURT
14	636.4	57.9	874	4	BI757540	BI757540 603029693
15	630.8	57.3	947	4	BI410783	BI410783 602963556
16	625.6	56.9	809	4	BG473023	BG473023 602514976
17	624	56.7	915	5	BQ936634	BQ936634 AGENCOURT
18	609.8	55.4	705	7	CN348925	CN348925 170005328
19	604.6	55.0	919	5	BU155142	BU155142 AGENCOURT
20	599.2	54.5	753	4	BG424731	BG424731 602453575
21	578.2	52.6	689	4	BI559635	BI559635 603252494
22	558.4	50.8	561	4	BM798461	BM798461 K-EST0081
23	542.4	49.3	760	5	BU055535	BU055535 UI-M-FOO-
24	541.8	49.3	545	7	CN348928	CN348928 170005313

```
Db 109 ATGCCCTGGGGCTCTTGGGAGACCGGCGGGTTACTGCTCTACTGCCCCCTGCATNGGGGGC 168
Qy 61 TTGTGGAAACAGGCGCATCATCTTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120
Db 169 TTGTGGAAACAGGCGCGATCATCTTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 228
Qy 121 ACCTTGGCTGTCCCTCTCTGGATTGAGTACAGCATCACAAAGCTCTCTTTACCAACCTC 180
Db 229 ACCTTGGCTGTCCCTCTCTGGATTGAGTACAGCATCACAAAGCTCTCTTTACCAACCTC 288
Qy 181 CATGTGCTTACCAAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGTCATC 240
Db 289 CATGTGCTTACCAAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGTCATC 348
Qy 241 ACCTTGGAGGATTTATGAGAGAGCTGGGACCCACCCACTTGGCCCCCTGAGAACGGGGT 300
Db 349 ACCTTGGAGGATTTATGAGAGAGCTGGGACCCACCCACTTGGCCCCCTGAGAACGGGGT 408
Qy 301 GCATACCTGCTTTGAGGTGGCAGCCAGCGAAGCCAGATGAAGAACGCTGCCCATGAAG 360
Db 409 GCATACCTGCTTTGAGGTGGCAGCCAGCGAAGCCAGATGAAGAACGCTGCCCATGAAG 468
Qy 361 GAAGGAACCCCTTTGGCCCATTTCTGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
Db 469 GAAGGAACCCCTTTGGCCCATTTCTGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 528
Qy 421 CTTTTCAGGCGATTTCTTCTAGTCTTCTTACAGAAACAATGGAGCAGAGATTTCT 480
Db 529 CTTTTCAGGCGATTTCTTCTAGTCTTCTTACAGAAACAATGGAGCAGAGATTTCT 588
Qy 481 CCAAGGAACATCGGTGCTTGGCTTGCCTGCGAGGCCCCAGTTCCTCCCTGCTAGAA 540
Db 589 CCAAGGAACATCGGTGCTTGGCTTGCCTGCGAGGCCCCAGTTCCTCCCTGCTAGAG 648
Qy 541 GAAACAGGCGCATACAGAGTACATGTTATGGTTCAGACGAATGGTGAAGACGGGAGAG 600
Db 649 GAAACAGGCGCATACAGAGTACATGTTATGGTTCAGACGAATGGTGAAGACGGGAGAG 708
Qy 601 GCCAGATTCATGCGCCACTTGTGCGGCCCTATGTGGGCAATTCATCTGGCATTTGGCTCT 660
Db 709 GCCAGATTCATGCGCCACTTGTGCGGCCCTATGTGGGCAATTCATCTGGCATTTGGCTCT 768
Qy 661 GACTGGAAGAACGCTGTGTCATGCTGAAGACGGGACTGCGAGGCTCGCACTTCATGCGC 720
Db 769 GACTGGAAGAACGCTGTGTCATGCTGAAGACGGGACTGCGAGGCTCGCACTTCATGCGC 828
Qy 721 TCTCCGCAAGTGTGGGCTACAGCGCGACAGCGGCCCTCCACGATGACTATGTGC 780
Db 829 TCTCCGCAAGTGTGGGCTACAGCGCGACAGCGGCCCTCCACGATGACTATGTGC 888
Qy 781 CTGCTGACCTGAAGGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGGTGCGTGGATGCC 840
Db 889 CTGCTGACCTGAAGGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGGTGCGTGGATGCC 948
Qy 841 CAGTCGGTCTACGTTGCTACTGATTCAGAGAGTTATGTGCTGAGCTCCACAGGCTCTC 900
Db 949 CAGTCGGTCTACGTTGCTACTGATTCAGAGAGTTATGTGCTGAGCTCCACAGGCTCTC 1008
Qy 901 AAAGGAAGGTGAAGGTGAGCTGAAGCGCTGAGGTGGCGCAGGTGCACTGTGATCATC 960
Db 1009 AAAGGAAGGTGAAGGTGAGCTGAAGCGCTGAGGTGGCGCAGGTGCACTGTGATCATC 1068
Qy 961 CTCGGCCAAAGCCGACCATTTATTTGSCAACTGTGTCTCTCTCTCTCTCTCTCTCTCTCT 1020
Db 1069 CTCGGCCAAAGCCGACCATTTATTTGSCAACTGTGTCTCTCTCTCTCTCTCTCTCTCTCT 1128
Qy 1021 CGGGAGCGGACCTCCAGGGAGGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1080
Db 1129 CGGGAGCGGACCTCCAGGGAGGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1188
Qy 1081 CTGCGGAGCGAGTTCTGATT 1100
Db 1189 CTGCGGAGCGAGTTCTGATT 1208
```

RESULT 2  
BX353274  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

BX353274 1141 bp mRNA linear EST 08-APR-2004  
BX353274 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens  
cDNA clone CS0DC003YJ11 5-PRIME, mRNA sequence.

BX353274  
EST.  
BX353274.2 GI:46290116

Homo sapiens (human)

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1. (bases 1 to 1141)

Li, W.B., Gruber, C., Jesse, J. and Polayes, D.

Full-length cDNA libraries and normalization

Unpublished (2001)

On May 5, 2003 this sequence version replaced gi:30371746.

Contact: Genoscope

Genoscope - Centre National de Sequencage

2 rue Gaston Cremieux, CP 5706 - 91057 Evry cedex - FRANCE

Email: [segref@genoscope.cns.fr](mailto:segref@genoscope.cns.fr), Web: [www.genoscope.cns.fr](http://www.genoscope.cns.fr)

1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime

end enriched, double-strand cDNA was digested with Not I and cloned

into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library

was normalized. Library was constructed by Life Technologies, a

division of Invitrogen. This sequence belongs to sequence cluster

9980.f

For more information about this cluster, see

<http://www.genoscope.cns.fr/cdna?s=CS0DC003CS060P1&c=9980.f>.

Location/Qualifiers

1. 1141

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="CS0DC003YJ11"

/tissue\_type="NEUROBLASTOMA COT 25-NORMALIZED"

/clone\_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"

/notes="1st strand cDNA was primed with a NotI-oligo(dT)

primer. Five prime end enriched, double-strand cDNA was

digested with Not I and cloned into the Not I and EcoR V

sites of the pCMVSPORT 6 vector. Library was normalized."

# ORIGIN

Query Match 81.2%; Score 892.8; DB 5; Length 1141;  
Best Local Similarity 92.8%; Pred. No. 8.9e-235;  
Matches 952; Conservative 32; Mismatches 35; Indels 7; Gaps 5;

Qy 1 ATGCCCGGGGCTCTTGGGACCGGCGGGTTACCTGCTCTACTGCCCCCTGCATGGGGGC 60

Db 86 ATGCCCGGCGCTCTTGGGACCGGCGGGTTACCTGCTCTACTGCMCTGCATGGGGGC 145

Qy 61 TTGTGGACACAGGCGCATCATCTTTGGGCTCTCTGGCATTTGCAAGCTGCTAAACCGT 120

Db 146 TTGTGGAAACAGGCGCATTCTTTGGGCTCTCTGGCATTTGCAAGCTGCAAAAMGT 205

Qy 121 ACCTTGGCTGTCCCTCTTGGATTGATACAGCATCACAAAGCTCTCTTCCACCACTC 180

Db 206 ACCTTGGCTGTACCTCTCTTGGATTGATACAGCATCAMAAGMTCTCTTCCACCACTC 265

Qy 181 CATGTGCTCTACCAAGATGACTTCAAGCTGGAGCCCTCCAGGCTTACCATGGGTCATC 240

Db 266 CATGTGCTCTACCAAGATGACTTMAAAMGGAGCACCTTMCAGGCTTACMATGGGTCATC 325

Qy 241 AGCTTGGAGGATTTTATGGAAGAGCTGGCACCCACCTCCAGGCTTCCAGAGCGGGT 300

Db 326 AGCTTGGAGGATTTTATGGAAGAGCTGGCACCCACCTCCAGGCTTCCAGAGCGGGT 385

Qy 301 GCATACCTGTTGAGGTGGCAGCCAGGCGAGGCTCAGATAGAGAGCTGCCCATGAAG 360

Db 386 GCATACCTGTTGAGGTGGCAGCCAGGCGAGGCTCAGATAGAGAGCTGCCCATGAAG 444

361	Qy	GAAGGAACCCCTTTGGCCCAATCTGGGATCAGTTTTCATGTAGTTTCAACAAGTCGGAG	420
445	Db	GAAGGAACCCCTTTGGCCCAATCTGGGATCAGTTTTCATGTAGTTTCAACAAGTCGGAG	504
421	Qy	CTTTTACAGGCATTTCTCCTTCAGTGTCTTCTACAGAGAAACAATGGAGCCAGAGATTTTCT	480
505	Db	CTTTTACAGGCATTTTCTTCAGTGTCTTCTACAGAGAAACAATGGAGCCAGAGATTTTCT	564
481	Qy	CCAAAGGAAACATCCGGTGTCTTCCCTGCCAGGAGCCCAAGTTCCTCCCTCTAGAA	540
565	Db	CCAAAGGAAACATCCGGTGTCTTCCCTGCCAGGAGCCCAAGTTCCTCCCTCTAGAG	624
541	Qy	GAACACAGGCCACTACAGAGAGTACATGGTATGGTTCAGACGAAATGGTGAAGACGGGAGAR	600
625	Db	GAACACAGGCCACTACAGAGAGTACATGGTATGGTTCAGACGAAATGGTGAAGACGGGAGAR	684
601	Qy	GCCACAGATTTCATGCCCACTTGTCCGGCCCTTATGTGGGCATTCATCTGCCGATTTGGCTCT	660
685	Db	GCCACAGATTTCATGCCCACTTGTCCGGCCCTTATGTGGGCATTCATCTGCCGATTTGGCTCT	744
661	Qy	GACTGGGAAGAACGCTGTGCGCATCTGCAAGGACGGGACTCGACGCTCGCACTTCATGCGCC	720
745	Db	GACTGGGAAGAACGCTGTGCGCATCTGCAAGGACGGGACTCGACGCTCGCACTTCATGCGCC	804
721	Qy	TCTCCGAGTGTGTGGCTACAGCCGAGACAGCGGCCCTCTCAGATGACTATGTGC	780
805	Db	TCTCCGAGTGTGTGGCTACAGCCGAGACAGCGGCCCTCTCAGATGACTATGTGC	864
781	Qy	CTGCTGACCTGAAGAGAGATCCAGAGGGCTGTGAAGCTCTTGGTGAAGTTCGCTGGATGC	839
865	Db	CTGCTGACCTGAAGAGAGATCCAGAGGGCTGTGAAGCTCTTGGTGAAGTTCGCTGGATGC	924
840	Qy	CCAGTCCGCTCTACGTTTCTACTGATTCGAGAGAGTTATGTGCTCGAGCTCCAAACAGCTCTT	899
925	Db	CCAGTCCGCTCTACGTTTCTACTGATTCGAGAGAGTTATGTGCTCGAGCTCCAAACAGCTCTT	984
900	Qy	CAAAAGGAAAGGTGAAGTGTGTGAGCTTGAAGCCCTGAGTGAGTGCCCAAGTCTGTACAT	959
985	Db	CAAAAGGAAAGGTGAAGTGTGTGAG-CTKAAGCYTGAAGKGGSCC--AGGTGCACTGTATAT	1041
960	Qy	CCTCGGCCAAGCCGACCACTTATTGGCAACTGTGTCTCTCTCTTCACTGCCCTTGTGAA	1019
1042	Db	CCTCGGC--AAGCGRACACTTATTGGCAATKTKTTCTCTTATTCTTTGKAAGGGG	1099
1020	Qy	GCGGGA 1025	
1100	Db	AGGGGA 1105	
RESULT 3			
AK081059			
LOCUS			
DEFINITION			
AK081059 3611 bp mRNA linear HTC 03-APR-2004			
Mus musculus 10 days neonate cerebellum cDNA, RIKEN full-length			
enriched library, clone:B930076G19 product:protein			
O-fucosyltransferase 1, full insert sequence.			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
Mus musculus (house mouse)			
Mus musculus			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus;			
REFERENCE			
1			
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,			
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.			
Normalization and subtraction of cap-trapper-selected cDNAs to			
prepare full-length cDNA libraries for rapid discovery of new genes			
MEDLINE			
PUBMED			
10349636			
REFERENCE			
2			
Carninci, P. and Hayashizaki, Y.			
High-efficiency full-length cDNA cloning			
Meth. Enzymol. 303, 19-44 (1999)			
AUTHORS			
TITLE			





## JOURNAL

Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.jp, URL: <http://genome.gsc.riken.jp/>; Tel: 81-45-503-9222, Fax: 81-45-503-9216)

## COMMENT

cdNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Retina RNA was provided by Dr. Stefano Gustincich (Department of Neurobiology, Harvard Medical School, 220 Longwood Ave., Boston, MA 02115, USA) whose assistance is gratefully acknowledged. Please visit our web site for further details.

URL: <http://genome.gsc.riken.jp/>  
URL: <http://fantom.gsc.riken.jp/>

FEATURES  
source

Location/Qualifiers  
1..2616  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="FANTOM\_DB:A930028F21"  
/db\_xref="taxon:10090"  
/clone="A930028F21"  
/tissue\_type="retina"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="adult"  
34..1215

## CDS

/note="unnamed protein product; protein  
O-fucosyltransferase 1 (MGD|MGI:2153207, GB|NM\_080463,  
evidence: BLASTN, 99%, match=1290)  
putative"

## ORIGIN

Query Match 80.4%; Score 884.2; DB 3; Length 2616;  
Best Local Similarity 87.9%; Pred. No. 2.6e-232;  
Matches 964; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

QY 3 GCCCGCGGCTCTGGACCGCGGTTACCTGCTCTACTGCTCCCTCGATGGGCGCTT 62  
DB 120 GTACGCGGCTCTGGACCGCTGGCGGTTACCTGCTCTACTGCTCCCTCGATGGGCGCTT 179

QY 63 TGGGAACAGCGCGATCTCTGGGCTCTGCGCATTTGCAAGCTGCTAAACCGTAC 122  
DB 180 TGGGAACAGCGCTGATCTCTGGGCTCTGCGCATTTGGAAGCTGCTGAACCGCAC 239

QY 123 CTGCTGCTCCCTCTGGATTGAGTACGAGCATCAAGAGCTCTCTTTCACCACTCCA 182  
DB 240 CTGCTGCTACCTCATGGATTGATACCAATCAAGCTCTCTTTCACCACTCCA 299

QY 183 TGTGCTTACCAAGAGTACTTCAAGCTGAGCGCCCTCCAGGCTTACCATCGGGTCATCAG 242  
DB 300 TGTGCTTACCAAGAGTACTTCAAGCTGAGCGCCCTCCAGGCTTACCATCGGGTCATCAG 359

QY 243 CTGAGGAGTTCATGGAGAGCTGGACCCACCTGCGCCCTGAGAGCGGGTGGC 302  
DB 360 CTGAGGAGTTCATGGAGAGCTGGACCCACCTGCGCCCTGAGAGCGGGTGGC 419

QY 303 ATACTGCTTTGAGTGGGAGCGGACCGAGCCAGATAAGAGAGCTGCCCATGAAGA 362

DB 420 ATACTGCTTTGAGTGGGAGCGCCAGCAAGTCTCTGATAAGAAGACATGTCCCATGAAGA 479

QY 363 AGGAAACCCCTTTGGCCCATTTCTGGATCAGTTCATGTGAGTTTCAACAAGTCGGAGCT 422  
DB 480 AGGAATCCTTTGGGCCATTTCTGGGACAGTTTCTGAGTTTCAATAGTCAAGACT 539

QY 423 TTTTACAGGCATTTCTTTCAGTGTCTCTACAGAGAACAAATGGAGCCAGAGATTTTCTCC 482  
DB 540 GTTACAGGCATTTCTTTCAGCGCTCTCTACAAAGAACAAATGGAGCCAGAGATTTTCTCC 599

QY 483 AAAGAAACATCGGTGCTTGGCCCTGCGAGGAGCCCGCCAGTTCCTCCCTCTCTAAGA 542  
DB 600 AAAAGAGCATCTGTGCTCGCACTGCTGGGGCCCGCCAGCAGTTCCTGTCTCGAGGA 659

QY 543 ACACAGGCCACTACAGAGTACATGATGTGTGTCAGAACAAATGGTGAAGCGGAGAGCG 602  
DB 660 ACACAGGAGCTCCAGAAATGATGTGTGTGATGAGATGGTGAAGCGGAGAGCG 719

QY 603 CCAGATTCATGCCCACTTGTTCGGCCCTATGTGGGCATTTCTGCGCATTTGGCTCTGA 662  
DB 720 CCTGATCAGTTCCCACTCGTCCGCGCTATGTGGGCATTTCTGCGCATTTGGCTCTGA 779

QY 663 CTGGAAGAACCTGTGTCATGTGAAGACGGAGCTGAGGCTCGCATTCATGCGCTC 722  
DB 780 CTGGAAGAAATGCTGTGTCATGTGAAGATGGAATCTGAGGGTCACTTTCATGGCTTC 839

QY 723 TCCGAGTGTGTGGGTACAGCCGACGACGCGCCCTCTGAGTGTGCTGAGTGTGCTGCT 782  
DB 840 CCTCAGTGTGTGGGTATAGCCGACGACGCGCCCTCTTCCATGATGATGCTGCT 899

QY 783 CCCTGACCTGAAGAGATCCAGAGGCTGTGCAAGCTCTGGGTGAGGTGCGTGGATGCCCA 842  
DB 900 CCCTGACCTGAAGAAATCCAGCGGCTGTGACGCTTTGGGTGAGAGCACTGAATGCCAG 959

QY 843 CTCGCTTACCTGTGCTACTGATTCGAGAGTTATGTGCTGAGTGTGCTGAGTGTGCTTCAA 902  
DB 960 ATCGGCTTACATCGCCACAGACTCTGAGAGCTAAGTGTGAGAGATCCAGCACTCTTCAA 1019

QY 903 AGGGAAGTGAAGTGTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 962  
DB 1020 AGAAGGTGAGGTGTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 1079

QY 963 CGGCAAGCCCACTTATTTGCAACTGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1022  
DB 1080 CGGCAAGCTGACCACTTATTTGCAACTGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1139

QY 1023 GGAGCGGACCTCCAGGGAGCGCT 1082  
DB 1140 GGAGCGGACCTCCAGGGAGCGCT 1199

QY 1083 CGGGAAGCTTCTGAT 1099  
DB 1200 TCGGATGAATTTGAT 1216

RESULT 5  
BM909055  
LOCUS  
DEFINITION  
AGENCOURT 6611915 NIH\_MGC\_98 Homo sapiens cDNA clone IMAGE:5431669  
5', mRNA sequence.  
ACCESSION  
BM909055  
VERSION  
BM909055.1 GI:19359434  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 1091)  
AUTHORS  
NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE  
National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL  
Unpublished (1999)  
COMMENT  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)



QY 61 TTGGGAACAGCCGATCACTTCTGGGCTCTGGCATTTGCAAGTCTGTAACCGT 120  
 Db 169 TTGGGAACAGCCGATCACTTCTGGGCTCTGGCATTTGCAAGTCTGTAACCGT 228  
 QY 121 ACCTTGGCTGCTCCCTTGGATTGAGTACAGATCAAGACCTCTTTTCAACCACTTC 180  
 Db 229 ACCTTGGCTGCTCCCTTGGATTGAGTACAGATCAAGACCTCTTTTCAACCACTTC 288  
 QY 181 CATGTGCTCTACAGAGTACTTCAAGCTGAGAGCCCTCCAGCTTACCATCGGTCATC 240  
 Db 289 CATGTGCTCTACAGAGTACTTCAAGCTGAGAGCCCTCCAGCTTACCATCGGTCATC 348  
 QY 241 AGCTTGGAGGATTTCATGAGAGTGGCACCCACCCACTGCGCCCTCAGAGCGGGTG 300  
 Db 349 AGCTTGGAGGATTTCATGAGAGTGGCACCCACCCACTGCGCCCTCAGAGCGGGTG 408  
 QY 301 GCATGACTGTTGAGTGGAGCCAGCCAGAGCCAGATAGAGAGAGCGTGCCTCATGAAG 360  
 Db 409 GCATGACTGTTGAGTGGAGCCAGCCAGAGCCAGATAGAGAGAGCGTGCCTCATGAAG 467  
 QY 361 GAAGGAACCCCTTTGGGCCATTCTGGGATCAGTTTTCATGTGAGTTTCAACAAGTCGGAG 420  
 Db 468 GAAGGAACCCCTTTGGGCCATTCTGGGATCAGTTTTCATGTGAGTTTCAACAAGTCGGAG 527  
 QY 421 CTTTTTACAGGCAATTCCTTCACTAGTCTTCTACAGAGAACATGAGGACAGATTTTCT 480  
 Db 528 CTTTTTACAGGCAATTCCTTCACTAGTCTTCTACAGAGAACATGAGGACAGATTTTCT 587  
 QY 481 CCAAGGAAACATCCGGTGTGCTCCCTGCGAGAGCCCGACAGTTCCTCGTCTCTAGAA 540  
 Db 588 CCAAGGAAACATCCGGTGTGCTCCCTGCGAGAGCCCGACAGTTCCTCGTCTCTAGAG 647  
 QY 541 GAAACAGGCCACTACAGAGTACATGTGTGCTGAGAGAACATGAGTGAAGCGGAGAG 600  
 Db 648 GAACAGGCCACTACAGAGTACATGTGTGCTGAGAGAACATGAGTGAAGCGGAGAG 707  
 QY 601 GCCCAGATTATGCCCACTTGTCCGGCCCTATGTGGGCAATTCATCTGCGCATTTGGCTCT 660  
 Db 708 GSCCAGATTATGCCCACTTGTCCGGCCCTATGTGGGCAATTCATCTGCGCATTTGGCTCT 767  
 QY 661 GACTGGAAGACGCTGTGCCATGCTGAAGAGCGGAGTGCAGGCTCGCACTTCATGCCC 720  
 Db 768 GACTGGAAGACGCTGTGCCATGCTGAAGAGCGGAGTGCAGGCTCGCACTTCATGCCC 827  
 QY 721 TCTCCGAGTGTGT 734  
 Db 828 TCTCCGAGTGTGT 841

RESULT 7  
 AL135434  
 LOCUS DKEZp76200115\_r1 762 (synonym: hm12) Homo sapiens cDNA clone  
 DEFINITION DKEZp76200115\_5', mRNA sequence.  
 ACCESSION AL135434  
 VERSION AL135434.1 GI:6603621  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 719)  
 Anorge,W., Wirkner,U., Meves,W., Weil,B. and Wiemann,S.  
 EST (Anorge,W., Wirkner,U., Meves,W., Weil,B. and Wiemann,S.)  
 Unpublished (1999)  
 Contact: MIPS  
 MIPS  
 Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany  
 This is the 5' sequence of the clone insert  
 Clone from S. Wiemann, Molecular Genome Analysis, German Cancer  
 Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;  
 sequenced by EMBL (European Molecular Biology Laboratories,  
 Heidelberg/Germany) within the cDNA sequencing consortium of the

German Genome Project.  
 No sl sequence available.  
 This clone (DKFZp76200115) is available at the RZPD in Berlin.  
 Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059  
 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.  
 Location/Qualifiers  
 1..719  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="DKFZp76200115"  
 /tissue type="melanoma (MeWo cell line)"  
 /dev stage="adult"  
 /lab\_host="DH10B"  
 /clone\_lib="762 (synonym: hm12)"  
 /note="vector: pSport1; Site\_1: NotI; Site\_2: SalI"

FEATURES  
 source

ORIGIN

Query Match 65.2%; Score 717.4; DB 1; Length 719;  
 Best Local Similarity 99.9%; Pred. No. 1.9e-186;  
 Matches 718; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 360 GGAAGGAACCCCTTTGGCCCATTTCTGGATCAGTTTTCATGTGAGTTTCAACAAGTCGGA 419  
 Db 1 GGAAGGAACCCCTTTGGCCCATTTCTGGATCAGTTTTCATGTGAGTTTCAACAAGTCGGA 60  
 QY 420 GCTTTTACAGGCAATTCCTTCACTAGTCTTCTACAGAGAACATGAGGACAGATTTTCT 479  
 Db 61 GCTTTTACAGGCAATTCCTTCACTAGTCTTCTACAGAGAACATGAGGACAGATTTTCT 120  
 QY 480 TCCAAAGGAACATCCGGTGTGCTCCCTGCGAGAGCCCGACAGTTCCTCGTCTCTAGA 539  
 Db 121 TCCAAAGGAACATCCGGTGTGCTCCCTGCGAGAGCCCGACAGTTCCTCGTCTCTAGA 180  
 QY 540 AGAACAACAGGCCACTACAGAGTACATGTGTGCTGAGAGAACATGAGTGAAGCGGAGAG 599  
 Db 181 GGAACAACAGGCCACTACAGAGTACATGTGTGCTGAGAGAACATGAGTGAAGCGGAGAG 240  
 QY 600 GGCCAGATTATGCCCACTTGTCCGGCCCTATGTGGGCAATTCATCTGCGCATTTGGCTCT 659  
 Db 241 GGCCAGATTATGCCCACTTGTCCGGCCCTATGTGGGCAATTCATCTGCGCATTTGGCTCT 300  
 QY 660 TGACTGGAAGAACCCCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 719  
 Db 301 TGACTGGAAGAACCCCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360  
 QY 720 CTCTCCGAGTGTGTGGGCTACAGCCGAGACAGCGGCCCGCCCTCAGATGACTATGTG 779  
 Db 361 CTCTCCGAGTGTGTGGGCTACAGCCGAGACAGCGGCCCGCCCTCAGATGACTATGTG 420  
 QY 780 CTTGCTGACCTGAAGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGGTCTGCTGGATGC 839  
 Db 421 CTTGCTGACCTGAAGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGGTCTGCTGGATGC 480  
 QY 840 CCAGTCTGCTTACCTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 899  
 Db 481 CCAGTCTGCTTACCTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540  
 QY 900 CAAAGGAAGGTGAAGTGTGAGCTTGAAGCTTGAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 959  
 Db 541 CAAAGGAAGGTGAAGTGTGAGCTTGAAGCTTGAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 600  
 QY 960 CTTGCGCAAGCCGACCACTTTTATTTGCAACTGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1019  
 Db 601 CTTGCGCAAGCCGACCACTTTTATTTGCAACTGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 660  
 QY 1020 GCGGAGCGGACCTCCAGGGGAGCGCT 1078  
 Db 661 GCGGAGCGGACCTCCAGGGGAGCGCT 719

RESULT 8  
 BI524031



```
Query Match      64.2%; Score 706.6; DB 5; Length 898;
Best Local Similarity 97.8%; Pred. No. 2e-183;
Matches 744; Conservative 4; Mismatches 10; Indels 3; Gaps 3;

QY 1 ATGCCCGGGGCTCTGGGACCGCGGCTTACCTGCTCTACTGCTCCCTGCTGATGGGGCG 60
DB 139 ATGCCCGGGGCTCTGGGACCGCGGCTTACCTGCTCTACTGCTCCCTGCTGATGGGGCG 198

QY 61 TTTGGGAACAGCGCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAAGCTGCTAAACCGT 120
DB 199 TTTGGGAACAGCGCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAAGCTGCTAAACCGT 258

QY 121 ACCTTGGGCTGCTCTCTTGGATTTGAGTACCAAGCATCAAGGCTCTCTTTTCAACACCTC 180
DB 259 ACCTTGGGCTGCTCTCTTGGATTTGAGTACCAAGCATCAAGGCTCTCTTTTCAACACCTC 318

QY 181 CATGTGCTCTACAGAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGTCATC 240
DB 319 CATGTGCTCTACAGAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGTCATC 378

QY 241 AGCTTGGAGGATTTATGAGAGAGTGGGACCCACCTGCTGCGCCCTGAGAGCGGGTG 300
DB 379 AGCTTGGAGGATTTATGAGAGAGTGGGACCCACCTGCTGCGCCCTGAGAGCGGGTG 438

QY 301 GCATCTGCTTTGAGGTGGACGCCAGCGAAGCCAGATTAAGAGACAGTGCCCATGAAG 360
DB 439 GCATCTGCTTTGAGGTGGACGCCAGCGAAGCCAGATTAAGAGACAGTGCCCATGAAG 497

QY 361 GAAGGAACCCCTTTGGCCATCTTGGGATCTGGGATCAGTTCATGTGAGTTTCAAGTCGGAG 420
DB 498 GAAGGAACCCCTTTGGCCATCTTGGGATCAGTTCATGTGAGTTTCAAGTCGGAG 557

QY 421 CTTTTTTACAGGATTTCTTTCAGTCTCTTACAGAGAACATGAGGACGAGATTTTCT 480
DB 558 CTTTTTTACAGGATTTCTTTCAGTCTCTTACAGAGAACATGAGGACGAGATTTTCT 617

QY 481 CCAAAGGAACATCCGGTGTCTGCTCCAGGAGCCCGAGCCAGTTCCTCCCTCTAGAG 540
DB 618 CCAAAGGAACATCCGGTGTCTGCTCCAGGAGCCCGAGCCAGTTCCTCCCTCTAGAG 677

QY 541 GAACACAGGCGCACTACAGAGTACATGTATGGTCAGAGC- AATGGGTGAAGACGGGAGA 599
DB 678 GAACACAGGCGCACTACAGAGTACATGTATGGTCAGAGCAAAATGTTGAAGACGGGAGA 737

QY 600 GGCCCGAGATTATGCCCCATCTTGTCCGCTTGTGAGGACGGGACTGAGGCTCGCCTC 659
DB 738 GGCCCGAGATTATGCCCCATCTTGTCCGCTTGTGAGGACGGGACTGAGGCTCGCCTC 797

QY 660 TGACTTGAAGAACGCTGTGCAATGCTGAAGGACGGGACTGAGGCTCGCCTC 719
DB 798 TGACTTGAAGAACGCTGTGCAATGCTGAAGGACGGGACTGAGGCTCGCCTC 857

QY 720 CT-CTCCGAGTGTGGGCTACAGCGGACGACAGCGGCC 759
DB 858 CTYCTCCGAGTGTGGGCTACAGCGGACGACAGCGGCC 898

RESULT 10
CO774833 LOCUS
ILLUMIGEN MCQ 51950 Katze MNLV Macaca nemestrina cDNA clone
IBUW:29960 57 similar to Bases 7 to 852 highly similar to human
POFUT1 (Hs.178292), mRNA sequence.
CO774833
CO774833.1 GI:50976096
EST.
Macaca nemestrina (pig-tailed macaque)
Macaca nemestrina
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
Cercopithecinae; Macaca.
REFERENCE
1 (bases 1 to 942)
```

AUTHORS  
TITLE  
JOURNAL  
COMMENT

Katze, M.G., Thomas, M., Korth, M., Iadonato, S.P. and Magness, C.L.  
Large-scale Rhesus Macaque cDNA Sequencing  
Unpublished (2003)  
Contact: C. Magness  
Illumigen Biosciences Inc.  
2203 Airport Way S, Suite 450, Seattle, WA 98134, USA  
Tel: 2063780400  
Fax: 2063780408  
Email: cmagness@illumigen.com  
Sequenced on 2004.07.21. 611 020 bases. Library Preparation: Prof.  
Michael Katze Lab at University of Washington DNA Sequencing:  
Illumigen Biosciences Inc. For further information, see  
http://www.macaque.org

## PCR Primers

FORWARD: CCCTCACTAAAGGGAACAAA

BACKWARD: CACTATAGGCGAATTGGGTA

Insert Length: 942 Std Error: 0.00

Plate: CL000531 row: E column: 09

Seq primer: CCCTCACTAAAGGGAACAAA

POLYA=No.

## Location/Qualifiers

1. 942

/organism="Macaca nemestrina"

/mol\_type="mRNA"

/strain="Indian"

/db\_xref="taxon:9545"

/clone="IBUW:29960"

/sex="male"

/lab\_host="Electronax DH10B"

/clone\_lib="Katze MNLV"

/note="Organ: liver; Vector: pDONR 222; Site 1: BsrG I;

Site 2: BsrG I; Created from Cloneminer cDNA Library

Construction kit (catalog #18249-029)"

## ORIGIN

Query Match 61.2%; Score 673.2; DB 7; Length 942;  
Best Local Similarity 94.1%; Pred. No. 3.3e-174;  
Matches 743; Conservative 0; Mismatches 43; Indels 4; Gaps 4;

QY 1 ATGCCCGGGGCTCTGGGACCGCGGTTACCTGCTCTACTGCTCCCTGCTGATGGGGCG 60  
DB 95 ATGCCCGGGGCTCTGGGACCGCGGTTACCTGCTCTACTGCTCCCTGCTGATGGGGCG 154  
QY 61 TTTGGGAACAGCGCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAAGCTGCTAAACCGT 120  
DB 155 TTTGGGAACAGCGCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAAGCTGCTAAACCG 214  
QY 121 ACCTTGGGCTGCTCTCTTGGATTTGAGTACCAAGCATCAAGGCTCTCTTTCAACACCTC 180  
DB 215 ACCTTGGGCTGCTCTCTTGGATTTGAGTACCAAGCATCAAGGCTCTCTTTCAACACCTC 274  
QY 181 CATGTGCTCTACAGAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGTCATC 240  
DB 275 CATGTGCTCTACAGAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGTCATC 334  
QY 241 AGCTTGGAGGATTTATGAGAGAGCTGGACCCACCTGAGAGCGGGTG 300  
DB 335 AGCTTGGAGGATTTATGAGAGAGCTGGACCCACCTGAGAGCGGGTG 394  
QY 301 GCATCTGCTTTGAGGTGGACGCCAGGAGCCAGATAAGAGACGCTGCCCATGAAG 360  
DB 395 GCATCTGCTTTGAGGTGGACGCCAGGAGCCAGATAAGAGACGCTGCCCATGAAG 454  
QY 361 GAAGGAACCCCTTTGGCCATCTTGGGATCAGTTCATGTGAGTTTCAACAGTCGGAG 420  
DB 455 GAAGGAACCCCTTTGGCCATCTTGGGATCAGTTCATGTGAGTTTCAACAGTCGGAG 514  
QY 421 CTTTTTTACAGGATTTCTTTCAGTCTCTTACAGAGAACATGAGGACGAGATTTTCT 480  
DB 515 CTTTTTTACAGGATTTCTTTCAGTCTCTTACAGAGAACATGAGGACGAGATTTTCT 574  
QY 481 CCAAAGGAACATCCGGTGTCTTGGCTTGCAGGAGCCCGAGCCAGTTCCTCCCTCTAGAA 540

Db	575	CCAAAGGNAACATCGGTGCTTGCCTCGCGGAGCCCGGCCCAAGTTCCCTGTCCTGGAA	634
Qy	541	GAACACAGGCGCACTACAGAGTACATGTATGCTCAGACGAATGCTGAAGACGGGAGG	600
Db	635	GAACACAGACCACTGCAAAAGTACATGTATGCTCAGACCAATGCTGAAGACGGGAGAA	694
Qy	601	GCCAGATTTCATGCCACCTTCT-CGGGCCCTATGTGGGCAATTCATCTCGCATTTGGC-T	658
Db	695	GCCAGAAATCCCGCCCACTTGTCCGGCCCTATGTGGGCAATTCCTCTAGCTAATGGCTT	754
Qy	659	CTGACTGGAAGAGCGCTGTGCCATGCTGGAAGACGGGACTGCAGCTCGCACTTCATGG	718
Db	755	CTGACTGGAAGAAATGCTGTGCTATGCTAAGGACGGAACTGSCAGCTCGCACTTCTTTGG	814
Qy	719	CCTCTCCGAGTGTGGG-CTACAGCCGACGACACAG-CGGCCCCCTCAGCATGACTAT	776
Db	815	CCTCTCCGAAATGTGGGGCCCTACGCCCCCAACACACAGCGCCGCCCTCCCGAAGACTAT	874
Qy	777	GTGCTGCTCT 786	
Db	875	GTGCTGCTCT 884	
RESULT 11			
LOCUS	BG747760		
DEFINITION	BG747760 808 bp mRNA linear EST 15-MAY-2001		
VERSION	602705255F1 NIH_MGC_43 Homo sapiens cDNA clone IMAGE:4842061.5,		
KEYWORDS	mRNA sequence.		
SOURCE	BG747760.1 GI:14058413		
ORGANISM	Homo sapiens (human)		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
JOURNAL	NIH-MGC <a href="http://mgi.nci.nih.gov/">http://mgi.nci.nih.gov/</a>		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)		
	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: <a href="mailto:cgapbs-t@mail.nih.gov">cgapbs-t@mail.nih.gov</a>		
	Tissue Procurement: AFCC		
	cDNA Library Preparation: Ling Hong/Rubin Laboratory		
	DNA Sequencing by: Incyte Genomics, Inc.		
	Clone distribution: MGC clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LLNL at:		
	<a href="http://image.llnl.gov">http://image.llnl.gov</a>		
	Plate: L1CMI675 row: n column: 14		
	High quality sequence stop: 808.		
FEATURES	Location/Qualifiers		
source	1..808		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:4842061"		
	/tissue_type="normal pigmented retinal epithelium"		
	/lab_host="DH10B (phage-resistant)"		
	/clone_lib="NIH MGC 43"		
	/note="Organ: eye; Vector: pOTB7; Site: 1: XhoI; Site 2:		
	EcoRI; cDNA made by oligo-dT priming. Directionally-		
	cloned into EcoRI/XhoI sites using the following 5'		
	adaptor: GGCAACGAG(G). Library constructed by Ling Hong		
	in the laboratory of Gerald M. Rubin (University of		
	California, Berkeley) using ZAP-cDNA synthesis kit		
	(Stratagene) and Superscript II RT (Life Technologies).		
	Note: this is a NIH_MGC Library.  "		
ORIGIN			
Query Match	60.28;	Score 662.2;	DB 4; Length 808;
Best Local Similarity	95.33;	Pred. No. 3.4e-171;	
Matches 693;	Conservative	0; Mismatches 33;	Indels 1; Gaps 1;

Qy	1	ATGCCCGGGCTCTCGGAGCCGGCGGTTACCTGTCTACTGCCCCCTGCATGGGGCGC	60
Db	82	ATGCCCGGGCTCTCGGAGCCGGCGGTTACCTGTCTACTGCCCCCTGCATGGGGCGC	141
Qy	61	TTTGGGAACGAGCGCATCACTTCTTGGGCTCTCTGGCAATTTGCAAGCTGTAAACCGT	120
Db	142	TTTGGGAACGAGCGCGATCACTTCTTGGGCTCTCTGGCAATTTGCAAGCTGTAAACCGT	201
Qy	121	ACCTTGGGCTGCTCTCTTGGATTTGAGTACAGGATCACAGCTCTCTTTCACCAACCTC	180
Db	202	ACCTTGGGCTGCTCTCTTGGATTTGAGTACAGGATCACAGCTCTCTTTCACCAACCTC	261
Qy	181	CATGTGCTCTACAGAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGTCATC	240
Db	262	CATGTGCTCTACAGAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGTCATC	321
Qy	241	AGCTTGGAGGATTTTCATGGAGAGCTTGGCACCCACCTGGCCCCCTGAGAGCGGGTG	300
Db	322	AGCTTGGAGGATTTTCATGGAGAGCTTGGCACCCACCTGGCCCCCTGAGAGCGGGTG	381
Qy	301	GCATATCTCTTTGAGGTGGCGAGCCAGCGAGCCAGATAGAGAGCGTCCCCCATGAAG	360
Db	382	GCATATCTCTTTGAGGTGGCGAGCCAGCGAGCCAGATAGAGAGCGTCCCCCATGAAG	441
Qy	361	GAAGAAACCCCTTTGGCCCATTTGGGATCAGTTTCAATGTGAGTTTCAACAGTCGGAG	420
Db	442	GAAGAAACCCCTTTGGCCCATTTGGGATCAGTTTCAATGTGAGTTTCAACAGTCGGAG	501
Qy	421	CTTTTTCACAGGCATTTCTTTCAGTGTCTTCTACAGAGAACTATGGAGCAGAGATTTTCT	480
Db	502	CTTTTTCACAGGCATTTCTTTCAGTGTCTTCTACAGAGAACTATGGAGCAGAGATTTTCT	561
Qy	481	CCAAAGGAACATCCGGTGTCTTGGCCCTGCCAGGAGCCCGCCAGTTTCCCGTCTAGAG	540
Db	562	CCAAAGGAACATCCGGTGTCTTGGCCCTGCCAGGAGCCCGCCAGTTTCCCGTCTAGAG	621
Qy	541	GAACACAGCCCACTACAGAGATACATGTGTATGTTTCAGACGAAATGGTGAAGAC-GGGAGA	599
Db	622	GAACACAGCCCACTACAGAGATACATGTGTATGTTTCAGACGAAATGGTGAAGACGGGAGA	681
Qy	600	GGCCAGAGATTCATGCCCACTTGTCCGGCCCTATGTGGGCAATTCATCTGGCATTTGGTCT	659
Db	682	GGCCAGAGATTCATGCCCACTTGTCCGGCCCTATGTGGGCAATTCATCTGGCATTTGGTCT	741
Qy	660	TGACTTGGAGAAACCGCTGTCCATGCTGAAGGACGGGACTGCGAGGCTCGCATTTATGGC	719
Db	742	TGACTTGGAGAAACCGCTGTCCATGCTGAAGGACGGGACTGCGAGGCTCGCATTTATGGC	801
Qy	720	CTCTCCG 726	
Db	802	CTCTCCG 808	
RESULT 12			
LOCUS	BP144264		
DEFINITION	BP144264 full-length enriched swine cDNA library, adult ovary Sus		
ACCESSION	scrofa cDNA clone OVR010034G09 5', mRNA sequence.		
VERSION	BP144264.1		
KEYWORDS	EST.		
SOURCE	Sus scrofa (pig)		
ORGANISM	Sus scrofa		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.		
	1 (bases 1 to 801)		
	Unishi, H., Eguchi, T., Suzuki, K., Sawazaki, T., Toki, D., Shinkai, H.,		
	Okumura, N., Hamasima, N. and Awata, T.		
	PEDE (Pig EST Data Explorer): construction of a database for ESTs		
	derived from porcine full-length cDNA libraries		
	Nucleic Acids Res. 32 (1), D484-D488 (2004)		
	Contact: Hirohide Unishi		
	Animal Genome Laboratory, Genome Research Department		
TITLE			
JOURNAL			
COMMENT			

National Institute of Agrobiological Sciences  
2 Ikenodai, Tsukuba, Ibaraki 305-8602, Japan.  
Tel: +81-29-838-8627  
Fax: +81-29-838-8627

Email: huenishia@affrc.go.jp

EST project with full-length enriched cDNA libraries carried out in  
Animal Genome Research Program (Japan) by National Institute of  
Agrobiological Sciences and STAFF-Institute  
Single pass sequencing of clones derived from oligo-capped cDNA  
library

Vector sequences were eliminated by RepeatMasker version 2002/07/13  
and crossmatch version 0.990319  
Low quality bases were trimmed based on the quality values.

## FEATURES

source

1. 801  
Location/Qualifiers  
/organism="Sus scrofa"  
/mol\_type="mRNA"  
/db\_xref="taxon:9823"  
/clone="OVR010034G09"  
/tissue type="ovary"  
/dev stage="adult"  
/clone\_lib="full-length enriched swine cDNA library, adult  
ovary"

## ORIGIN

Query Match 60.0%; Score 660.2; DB 5; Length 801;

Best Local Similarity 89.0%; Pred. No. 1.2e-170;  
Matches 713; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 194 AGAAGTACTTCAAGTGGAGCCCTCCAGGCTTACCATCGGTCTATCAGCTTGGAGGATT 253  
DB 1 AAAAGTACTTCAAGTGGAGCCCTCCAGGCTTACCATCGGTCTATCAGCTTGGAGGACT 60  
QY 254 TCATGGAGAAGTGGCACCACCCACTGGCCCCCTGAGAGGGGTGCATCTGCTTTG 313  
DB 61 TCATGGAGAAGTGGCACCACCCACTGGCCCCCTGAGAGGGGTGCATCTGCTTTG 120  
QY 314 AGGTGGCAGCCAGCGAAGCCAGATAGAGAGCTGCCCATGAAGGAAGAAACCCCT 373  
DB 121 AGGTGGCAGCAGCGAAGCCAGATAGAGAGCATGCCCATGAAGGAAGAAACCCCT 180  
QY 374 TTGGCCCATCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAGCTTTTACAGGCA 433  
DB 181 TTGGCCCATTTTGGGATCAGTTTCATGTGAGTTTCAACAAGTCAGAACTTTTTCAGGCA 240  
QY 434 TTTCCTTCAGTGTCTTCTACAGAGAAATGAGAGCCAGAGATTTTCTCAAGGAACATC 493  
DB 241 TTTCCTTCAGTGTCTTCTACAGAGAAATGAGAGCCAGTGGATCCAAAGATTTTCTCAAGGAACATC 300  
QY 494 CGGTGCTTGGCTGCGCAGAGCCCGCAGCCAGTTCCTCCGCTCTAGAGAAACACAGGCCAC 553  
DB 301 CAGTGTCTTGGCTGCGCAGAGCCCGCAGTTCCTCCGCTCTAGAGAAACATAGGCCAC 360  
QY 554 TACAGAAGTACATGTATGTCAGACGAAATGGTGAAGACGGGAGGCCAGATTCATG 613  
DB 361 TTCAGAATACATGTGTGTCAGACGAGATGGTGAAGACGGGAGGCCAGATCCGTG 420  
QY 614 CCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGGCGCATTTGCTGTGACCTGGAAGAACG 673  
DB 421 CCCACCTTATCCGGCCCTATGTGGGCATTCATCTGGCGCATTTGCTGTGACCTGGAAGAACG 480  
QY 674 CTTGTGCCATCTGAAGAACGGGACTGCAGGCTCGACCTTCATGCTCTCCGAGTGTG 733  
DB 481 CGTGGCCATCTGAAGAACGGGACTGCAGGCTCGACCTTCATGCTCTCCGAGTGTG 540  
QY 734 TGGGCTACAGCGGACGAGCGGCCCTCCATGATGATGTGCTGCTGACCTGA 793  
DB 541 TGGGTTACAGCGGACGAGCGGCCCTCCATGATGATGTGCTCTCCGAGTGTG 600  
QY 794 AGGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGGTGCTGGATGCCAGTCTGCTACG 853  
DB 601 AGGAGATCAGCGCGCCCTGAAGCTCTGGGTGAGGTGCTGGATGCCAGTCTGCTTACA 660

QY 854 TTGCTACTGATTCCGAGAGTTTATGTGCTGAGCTCCAAAGCTCTTCAAAGGGAAGGTGA 913  
DB 661 TCGCCACGAGCTCTGAGAGTTTACCTGCGGAGATCCAGCAGCTCTTCAAAGGAAGGTGA 720  
QY 914 AGGTGGTGAAGCTGAAGCTGAGGTGGCCAGGTGACCTGTACATCTCGGCCAAGCG 973  
DB 721 AGGTGGTGAAGCTGAAGCTGAGGTGGCCAGGTGACCTGTACATCTTGGCCAGGTG 780  
QY 974 ACCACTTTATTTGGCAACTGTG 994  
DB 781 ACCACTTTATTTGGCAACTGTG 801

## RESULT 13

BO964179

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Mus musculus

(house mouse)

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1 (bases 1 to 988)

NIH-MSC http://nsl.nsl.nih.gov/

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: Dr. David Rowe

cDNA Library Preparation: Invitrogen Corp

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLNL4077 row: j column: 24

High quality sequence stop: 702.

Location/Qualifiers

1. 988

/organism="Mus musculus"

/mol\_type="mRNA"

/db\_xref="taxon:10090"

/clone="IMAGE:6509783"

/tissue type="undifferentiated limb"

/lab host="DH10B (phage-resistant)"

/clone\_lib="NIH\_MGC\_134"

/note="Vector: PCMV-SPOK6 1; Site 1: EcoRV; Site 2: NotI;

Cloned unidirectionally. Primer: Oligo dt. Average insert

size 1.7 Kb. Constructed by Resgen, Invitrogen Corp. Note:

this is a NIH\_MGC Library."

ORIGIN

Query Match

Best Local Similarity

Matches

Conservative

Mismatches

Indels

Gaps

2;

3

GCCC

CGGG

CGGG

CGGG

CGGG

CGGG

CGGG

CGGG

CGGG



Db 328 TGTGTCTCAACAAAGTACTCAAACTGGAGCCCTCCAAAGCCTACCATCGGGTTGTCAAG 387  
Qy 243 CTTGAGGATTTTCATGGAGAAGCTGGACCCACCCACTGGCCCCCTGTGAGAACGGGTGGC 302  
Db 388 CTTGAGGACTTTCATGAAAACTTGACACCCCTCCCACTGGCCCCCTGTGAGAACGGGTGGC 447  
Qy 303 ATACTGCTTTGAGTGGCAGCCAGCAGCCAGATTAAGAACAGCTGCCCATGAAGGA 362  
Db 448 ATACTGCTTTGAGTGGCAGCCAGCAGGAGTCTGTATGAAGACATGTCCCATGAAGGA 507  
Qy 363 AGGAACCCCTTTGGGCGCATTTGGGATCAGTTTCATGTGAGTTTCACAAAGTGGAGCT 422  
Db 508 AGGAATACCTTTTGGGCGCATTTGGGACAGTTTCATGTGAGTTTCATTAAGTCAAGACT 567  
Qy 423 TTTTACAGGCATTTCTCTCAGTGTCTTCTACAGAGAACAAATGGAGCCAGATTTTCTCC 482  
Db 568 GTTCACAGGCATTTCTCTCAGGCGCTCTCAAAAGAACAAATGGAGCCAGATTTCTCTGC 627  
Qy 483 AAAGAGACATCCGGTGTCTTCCCTGCCAGGAGCCAGCCAGCTTCCCGTCTCTAGAAGA 542  
Db 628 AAAGAGACATCTGTGTCTGCACTGCCTGNGGCCCCAGCACAGTTCCCTGTCTTGAGGA 687  
Qy 543 ACAGAGCCACTACAGAGTACATGTATGTGTGACAGCAATGGTGAAGAGGAGGC 602  
Db 688 ACACAGGAGCTTCAGAGTACATGTGTGTGATGATGATGATGATGATGATGATGATGATG 747  
Qy 603 CCAGATTCATGCCACCTTGTTCGGCCCTATGTGGGCAATCATCTGCGCATTTGGCTCTGA 662  
Db 748 CTTGATCAGTGGCCACCTCTGTCGGCCCTATGTGGGCAATCATCTGCGCATTTGGCTCTGA 807  
Qy 663 CTGGAAGAACGCTGTGCCATGTGCAAGGAGCGGACATGCAAGGCTGCACATTCATGGGCTC 722  
Db 808 CTGGAAGAACGCTGTGCCATGTGCAAGGATGGAATGCAAGGCTGCACATTCATGGGCTC 867  
Qy 723 TCC-GCAGTGTGGGTACAGCCAGCAGCAGCGGCCCTCCAGATGATATGTGCC 781  
Db 868 CCGTTCNNYGTGTGGGTATAGCCGACAGCAGCGGCCCTCTCAC-ATGACCAATGTGCC 926  
Qy 782 TGCCTGACCTGAGGAGATTCAGAGGCTGTGAAGCTCTGGGTGAG 827  
Db 927 TCCCTGACCTGAAGAAATCAGCGGCTGGGGACCGTTTGGGGTG 972

RESULT 14  
BI757540  
LOCUS  
DEFINITION 603029693P1 NIH\_MGC\_114 Homo sapiens cDNA clone IMAGE:5199862 5',  
mRNA sequence.  
ACCESSION BI757540  
VERSION BI757540.1 GI:15749118  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 874)  
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs@mail.nih.gov

Tissue Procurement: Life Technologies, Inc.  
CDNA Library Prepared by: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
CDNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov

Plate: L14M1500 row: j column: 23  
High quality sequence stop: 798.

FEATURES  
1..874  
Location/Qualifiers  
/organism="Homo sapiens"

/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5199862"  
/lab\_host="DH10B"  
/clone\_lib="NIH\_MGC\_114"  
/notes="Organ: brain; Vector: pCMV-SPORT6; Site 1: NotI;  
Site 2: EcoRV (destroyed); RNA source anonymous pool of 6  
male brains age range 23-27 yo. Library is oligo-dT  
primed and directionally cloned. (EcoRV site is destroyed  
upon cloning). Average insert size 1.5 kb, insert size  
range 1-3 kb. Library is normalized and enriched for  
full-length clones and was constructed by C. Gruber  
(Invitrogen). Research Genetics tracking code 019. Note:  
this is a NIH\_MGC Library."

ORIGIN

Query Match 57.9%; Score 636.4; DB 4; Length 874;  
Best Local Similarity 96.1%; Pred. No. 4.5e-164;  
Matches 717; Conservative 0; Mismatches 21; Indels 8; Gaps 6;  
Qy 1 ATGCCCGGGCTCTGGGACCCGGCGGTACTCTACTGCCCTGCATGGGGCGC 60  
Db 109 ATGCCCGGGCTCTGGGACCCGGCGGTACTCTACTGCCCTGCATGGGGCGC 168  
Qy 61 TTTGGGAACACAGCCGATCATCTTTTGGGCTCTCTGGCATTTGCAAGCTGCTAAACC-G 119  
Db 169 TTTGGGAACACAGCCGATCATCTTTTGGGCTCTCTGGCATTTGCAAGCTGCTAAACCCTG 228  
Qy 120 TACCTTGGCTGTCCCTCTCTGGATTGAGTACCAGCATCAAGAGCCTCTCTTACCACACT 179  
Db 229 TACCTTGGCTGTCCCTCTCTGGATTGAGTACCAGCATCAAGAGCCTCTCTTACCACACT 288  
Qy 180 CCATGTGTCTTACCAGAACTTCTCAAGCTGGAGCCCTCCAGGCTTACCATCGGGTCAAT 239  
Db 289 CCATGTGTCTTACCAGAACTTCTCAAGCTGGAGCCCTCCAGGCTTACCATCGGGTCAAT 348  
Qy 240 CAGCTTGGAGATTTTCATGGAGAAGCTGGCAACCCACCTGGCCCCCTGAGAAGCGGT 299  
Db 349 CAGCTTGGAGATTTTCATGGAGAAGCTGGCAACCCACCTGGCCCCCTGAGAAGCGGT 408  
Qy 300 GGCATATCTTTGAGGTGGAGCCAGCAAGCCAGATAAGAGAGCTGCCCATGAA 359  
Db 409 GGCATATCTTTGAGGTGGAGCCAGCAAGCCAGATAAGAGAGCTGCCCATGAA 468  
Qy 360 GGAAGAAACCCCTTTTGGCCCAATTTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGA 419  
Db 469 GGAAGAAACCCCTTTTGGCCCAATTTCTGGGATCAG-TTCATGTGAGTTTCAACAAGTCGGA 527  
Qy 420 GCTTTTACAGGCATTTCTTTCAGTGTCTTCTACAGAGAACATATGGAGCCAGAGATTTTC 479  
Db 528 GCTTTTACAGGCATTTCTTTCAGTGTCTTCTACAGAGAACATATGGAGCCAGAGATTTTC 587  
Qy 480 TCCAAAGGAACATCGGTGTCTTGGCCCTGCGAGAGCCCT-AGCCAGTTCCCGTCTAG 538  
Db 588 TCCAAAGGAACATCGGTGTCTTGGCCCTGCGAGAGCCCTAGAGCCAGTTCCCGTCTAG 647  
Qy 539 AAGAACACAGGCCACTACAGAGTACATGTATGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 596  
Db 648 AAGAACACAGGCCACTACAGAGTACATGTATGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 707  
Qy 597 AGAGGCCAGATTCATGCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTTG 656  
Db 708 AGAGGCCAGATTCATGCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTTG 767  
Qy 657 CT-CTGACTGGAAGACCGCTGTGCGATCTGA--AGAGGGGACTGAGGCTGCGACTT 713  
Db 768 CTCCTGACTGAGAGACCGCTGTGCGATCTGACCGGACTGTGGGCTTCCCTTCCCTTC 827  
Qy 714 CAGTGGCTCTCCGAGTGTGTGGGT 739  
Db 828 ATGGCTCTCTCCGAGTGTGTGGGT 853





THIS PAGE BLANK (USPTO)

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 9, 2005, 08:00:04 ; Search time 984.698 Seconds  
(without alignments)  
9238.165 Million cell updates/sec

Title: US-09-774-954-16

Perfect score: 1100

Sequence: 1 ATGCCCGGGGCTCTGGGA.....CTGCGGAGGAGTTCTGATT 1100

Scoring table: IDENTITY NUC.

Gapop 10.0 , Gapext 1.0

Searched: 9794790 seqs, 413490567 residues

Total number of hits satisfying chosen parameters: 19589580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:\*

```
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US05_PUBCOMB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US04_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US03_PUBCOMB.seq.*
6: /cgn2_6/ptodata/1/pubpna/US02_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US01_PUBCOMB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US00_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US09D_PUBCOMB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US09E_PUBCOMB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US09F_PUBCOMB.seq.*
15: /cgn2_6/ptodata/1/pubpna/US09G_PUBCOMB.seq.*
16: /cgn2_6/ptodata/1/pubpna/US09H_PUBCOMB.seq.*
17: /cgn2_6/ptodata/1/pubpna/US09I_PUBCOMB.seq.*
18: /cgn2_6/ptodata/1/pubpna/US09J_PUBCOMB.seq.*
19: /cgn2_6/ptodata/1/pubpna/US09K_PUBCOMB.seq.*
20: /cgn2_6/ptodata/1/pubpna/US09L_PUBCOMB.seq.*
21: /cgn2_6/ptodata/1/pubpna/US09M_PUBCOMB.seq.*
22: /cgn2_6/ptodata/1/pubpna/US09N_PUBCOMB.seq.*
23: /cgn2_6/ptodata/1/pubpna/US09O_PUBCOMB.seq.*
24: /cgn2_6/ptodata/1/pubpna/US09P_PUBCOMB.seq.*
25: /cgn2_6/ptodata/1/pubpna/US09Q_PUBCOMB.seq.*
26: /cgn2_6/ptodata/1/pubpna/US09R_PUBCOMB.seq.*
27: /cgn2_6/ptodata/1/pubpna/US09S_PUBCOMB.seq.*
28: /cgn2_6/ptodata/1/pubpna/US09T_PUBCOMB.seq.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1100	100.0	1100	11	US-09-774-954-16
2	1100	100.0	1514	11	US-09-774-954-1
3	1098.4	99.9	1300	11	US-09-774-954-4
4	1098.4	99.9	11284	11	US-09-774-954-5
5	1096.8	99.7	5218	16	US-10-301-822-168

6	1096.8	99.7	5218	22	US-10-956-157-4752	Sequence 4752, Ap
7	1033.4	93.9	5009	11	US-09-774-954-7	Sequence 7, Appli
8	471.4	42.9	1400	22	US-10-956-157-9986	Sequence 9986, Ap
9	471.4	42.9	1506	22	US-10-956-157-4751	Sequence 4751, Ap
10	251.2	22.8	10331	10	US-09-764-891-6032	Sequence 6032, Ap
11	233	21.2	479	10	US-09-918-995-19437	Sequence 19437, A
12	221.8	20.2	479	10	US-09-918-995-22744	Sequence 22744, A
13	166.4	15.1	1209	26	US-11-097-143-18161	Sequence 18161, A
14	150.4	13.7	3264	26	US-11-097-143-18160	Sequence 18160, A
15	114.6	10.4	402	9	US-09-783-590-11501	Sequence 11501, A
16	108.6	9.9	3793	26	US-11-097-143-18082	Sequence 18082, A
17	71.8	6.5	3567	26	US-11-097-143-18169	Sequence 18169, A
18	60	5.5	60	10	US-09-908-975-17287	Sequence 17287, A
19	56	5.1	760	24	US-10-820-474A-236	Sequence 236, App
20	55.6	5.1	521	13	US-09-925-065A-60244	Sequence 60244, A
21	42.4	3.9	1438	18	US-10-282-122A-19431	Sequence 19431, A
22	41.2	3.7	114793	16	US-10-148-806-3	Sequence 3, Appli
23	41.2	3.7	114793	24	US-10-859-792-3	Sequence 3, Appli
24	40.4	3.7	1212	21	US-10-425-115-114097	Sequence 114097, A
25	40.4	3.7	5932	16	US-10-205-823-312	Sequence 312, App
26	40.4	3.7	5932	26	US-11-051-454-312	Sequence 312, App
27	40	3.6	40	11	US-09-774-954-10	Sequence 10, Appl
28	40	3.6	40	11	US-09-774-954-12	Sequence 12, Appl
29	40	3.6	2032	20	US-10-437-963-8098	Sequence 8098, Ap
30	39.2	3.6	618	13	US-09-925-065A-571171	Sequence 571171, A
31	39	3.5	372	18	US-10-282-122A-31476	Sequence 31476, A
32	38.6	3.5	407	9	US-09-960-352-11732	Sequence 11732, A
33	38.6	3.5	2358	10	US-09-919-039-190	Sequence 190, App
34	38.6	3.5	2384	22	US-10-278-698-123	Sequence 123, App
35	38.6	3.5	2384	22	US-10-278-698-638	Sequence 638, App
36	38.6	3.5	2670	17	US-10-247-671-64	Sequence 64, Appl
37	38.4	3.5	532	10	US-09-918-995-13517	Sequence 13517, A
38	38.4	3.5	591	16	US-10-156-761-3550	Sequence 3550, Ap
39	38.4	3.5	6227	19	US-10-240-425-364	Sequence 364, App
40	38.4	3.5	6227	22	US-10-278-698-100	Sequence 100, App
41	38.4	3.5	6227	22	US-10-278-698-615	Sequence 615, App
42	38.4	3.5	6256	22	US-10-278-698-101	Sequence 101, App
43	38.4	3.5	6256	22	US-10-278-698-616	Sequence 616, App
44	38.4	3.5	9025608	16	US-10-156-761-1	Sequence 1, Appli
45	37.8	3.4	1234	18	US-10-369-493-39272	Sequence 39272, A

#### ALIGNMENTS

#### RESULT 1

US-09-774-954-16  
; Sequence 16, Application US/09774954  
; Publication No. US20040241645A1  
; GENERAL INFORMATION:  
; APPLICANT: Yang Wang, Michael W. Spellman  
; TITLE OF INVENTION: O-Pucosyltransferase  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Genentech, Inc.  
; STREET: 1 DNA Way  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WinPatIn (Genentech)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/774,954  
; FILING DATE: 30-Jan-2001  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/978,741  
; FILING DATE: 26-NOV-1997  
; APPLICATION NUMBER: 08/792,498

```
;
; FILING DATE: 31-JAN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Svoboda, Craig G.
; REGISTRATION NUMBER: 39,044
; REFERENCE/DOCKET NUMBER: P1041P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-1489
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1100 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 16:
US-09-774-954-16

Query Match      100.0%; Score 1100; DB 11; Length 1100;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGCCCGCGGCTCTCTGGACCGCGCGGTACCTGCTCTACTGCGCCCTGATGGGCGC 60
Db 1 ATGCCCGCGGCTCTCTGGACCGCGCGGTACCTGCTCTACTGCGCCCTGATGGGCGC 60

Qy 61 TTGGGAACACGAGCGGATCACTTTTGGGCTCTCTGGCATTTGCCAAGCTGCTAAACCGT 120
Db 61 TTGGGAACACGAGCGGATCACTTTTGGGCTCTCTGGCATTTGCCAAGCTGCTAAACCGT 120

Qy 121 ACCTTGGGTGTCCTCTCTGGATTGAGTACAGCATCACAAAGCTCCTTTTACCAACCTC 180
Db 121 ACCTTGGGTGTCCTCTCTGGATTGAGTACAGCATCACAAAGCTCCTTTTACCAACCTC 180

Qy 181 CATGTGCTCTACAGAAAGTACTTCAAGCTGGAGCGCCCTCAGGCTTACCATCGGTCATC 240
Db 181 CATGTGCTCTACAGAAAGTACTTCAAGCTGGAGCGCCCTCAGGCTTACCATCGGTCATC 240

Qy 241 ACCTTGGAGGATTTTCATGGAGAGCTGGCACCCACCTGCGCCCTGAGAGCGGGTG 300
Db 241 ACCTTGGAGGATTTTCATGGAGAGCTGGCACCCACCTGCGCCCTGAGAGCGGGTG 300

Qy 301 GCNTACTGTTTGGGTGGCGAGCCGAGCGAGCCAGATAGAGAGCTGCCCCATGAAG 360
Db 301 GCNTACTGTTTGGGTGGCGAGCCGAGCGAGCCAGATAGAGAGCTGCCCCATGAAG 360

Qy 361 GAAGGAACCCCTTTGGCCCATTTCTGGATCAGTTTCATGTGAGTTTCAACAGTCGGAG 420
Db 361 GAAGGAACCCCTTTGGCCCATTTCTGGATCAGTTTCATGTGAGTTTCAACAGTCGGAG 420

Qy 421 CTTTTTACAGGCATTTCTTCAAGTCTTCTTACAGAGAACTATGGAGCCAGAGATTTTCT 480
Db 421 CTTTTTACAGGCATTTCTTCAAGTCTTCTTACAGAGAACTATGGAGCCAGAGATTTTCT 480

Qy 481 CCAAGGAACATCCCGTGTCTTGGCCCTGCCAGGAGCCCGAGTTCCTCCCTAGAA 540
Db 481 CCAAGGAACATCCCGTGTCTTGGCCCTGCCAGGAGCCCGAGTTCCTCCCTAGAA 540

Qy 541 GAACACAGCCCACTACAGAGTACATGGTATGTGACAGAAATGGTGAAGACGGAGAG 600
Db 541 GAACACAGCCCACTACAGAGTACATGGTATGTGACAGAAATGGTGAAGACGGAGAG 600

Qy 601 GCCAGATTTCAGCCCACTTGTCCGCGCTTGTGGGCTTTCATCTGCCATTGGCTCT 660
Db 601 GCCAGATTTCAGCCCACTTGTCCGCGCTTGTGGGCTTTCATCTGCCATTGGCTCT 660

Qy 661 GACTGGAAGACGCTGTGCTTGCATCAAGAGCGGACTGCGAGGCTGCACTTCATGGCC 720
Db 661 GACTGGAAGACGCTGTGCTTGCATCAAGAGCGGACTGCGAGGCTGCACTTCATGGCC 720

Qy 721 TCTCCGAGTGTGGGCTTACAGCGGAGCAGCGGCGCCCTCAGAGTATGTGC 780
Db 721 TCTCCGAGTGTGGGCTTACAGCGGAGCAGCGGCGCCCTCAGAGTATGTGC 780
```

```
Qy 781 CTGCTGCACCTGAAGGAGATCCAGAGGCTCTGAAGCTCTGGGTGAGTCTGCGTGAATGCC 840
Db 781 CTGCTGCACCTGAAGGAGATCCAGAGGCTCTGAAGCTCTGGGTGAGTCTGCGTGAATGCC 840

Qy 841 CAGTCGGTCTACGTTGCTACTACTGATTCGAGAGGTTATGTCCTGAGCTCCAAACAGCTCTTC 900
Db 841 CAGTCGGTCTACGTTGCTACTACTGATTCGAGAGGTTATGTCCTGAGCTCCAAACAGCTCTTC 900

Qy 901 AAAGGGAAGTGAAGTGTGAGCTGAAGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 960
Db 901 AAAGGGAAGTGAAGTGTGAGCTGAAGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 960

Qy 961 CTCGCCCAAGCGGACCACTTTATTTGGCACTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1020
Db 961 CTCGCCCAAGCGGACCACTTTATTTGGCACTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1020

Qy 1021 CGGAGAGCGGACCTCCAGAGGAGGCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1080
Db 1021 CGGAGAGCGGACCTCCAGAGGAGGCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1080

Qy 1081 CTGCGGGACGAGTTCTGATT 1100
Db 1081 CTGCGGGACGAGTTCTGATT 1100

RESULT 2
US-09-774-954-1
; Sequence 1, Application US/09774954
; Publication No. US20040241645A1
; GENERAL INFORMATION:
; APPLICANT: Yang Wang, Michael W. Spellman
; TITLE OF INVENTION: O-Pucosyltransferase
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/774,954
; FILING DATE: 30-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,741
; FILING DATE: 26-NOV-1997
; APPLICATION NUMBER: 08/792,498
; FILING DATE: 31-JAN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Svoboda, Craig G.
; REGISTRATION NUMBER: 39,044
; REFERENCE/DOCKET NUMBER: P1041P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-1489
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1514 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-774-954-1

Query Match      100.0%; Score 1100; DB 11; Length 1514;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy 1 ATGCCCCGGGGCTCTCTGGGACCCGGCCGGTTACCTGCTCTACTGTCCTCCCTGATGGGGCGC 60  
Db 1 ATGCCCCGGGGCTCTCTGGGACCCGGCCGGTTACCTGCTCTACTGTCCTCCCTGATGGGGCGC 60  
Qy 61 TTTGGGAACCAAGCCGATCACTTCTTGGGCTCTCTGGCAATTTGCAAGCTGCTAAACCGT 120  
Db 61 TTTGGGAACCAAGCCGATCACTTCTTGGGCTCTCTGGCAATTTGCAAGCTGCTAAACCGT 120  
Qy 121 ACCTTGGGCTGCTCCCTCTCTGGATTCAGTACAGCATCAAGCCCTCTTCCACCAACCTC 180  
Db 121 ACCTTGGGCTGCTCCCTCTCTGGATTCAGTACAGCATCAAGCCCTCTTCCACCAACCTC 180  
Qy 181 CATGTGCTCTACCAAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGCTCATC 240  
Db 181 CATGTGCTCTACCAAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGCTCATC 240  
Qy 241 AGCTTGGAGGATTTATGAGAGAGCTGGAGCCACCCACTGCGCCCTTGAGAAGCGGGTG 300  
Db 241 AGCTTGGAGGATTTATGAGAGAGCTGGAGCCACCCACTGCGCCCTTGAGAAGCGGGTG 300  
Qy 301 GCATACTCTTTGAGGTGGAGCCAGAGAGCCAGATAGAGAGCTGCGCCCATGAAG 360  
Db 301 GCATACTCTTTGAGGTGGAGCCAGAGAGCCAGATAGAGAGCTGCGCCCATGAAG 360  
Qy 361 GAAGGAACCCCTTTGGCCCATTTCTGGGATCAGTTTCTAGTGTGAGTTTCAACAAGTCGGAG 420  
Db 361 GAAGGAACCCCTTTGGCCCATTTCTGGGATCAGTTTCTAGTGTGAGTTTCAACAAGTCGGAG 420  
Qy 421 CTTTTTACAGGATTTCTTCACTGCTTCTACAGAGAAATGAGGACGAGATTTTCT 480  
Db 421 CTTTTTACAGGATTTCTTCACTGCTTCTACAGAGAAATGAGGACGAGATTTTCT 480  
Qy 481 CCAAGGAACATCGGTGCTGCTGCGCCAGAGAGCCAGCCAGCTTCCCGTCTCTAGAA 540  
Db 481 CCAAGGAACATCGGTGCTGCTGCGCCAGAGAGCCAGCCAGCTTCCCGTCTCTAGAA 540  
Qy 541 GAACACAGGCCACTACAGAAGTACATGTATGTGTGAGAGAAATGAGTGAAGACGGAGAG 600  
Db 541 GAACACAGGCCACTACAGAAGTACATGTATGTGTGAGAGAAATGAGTGAAGACGGAGAG 600  
Qy 601 GCCAGATTCATGCCCACCTTGTGCGGCCCTATGTGGGCATTCATCTGCGCATTTGGCTCT 660  
Db 601 GCCAGATTCATGCCCACCTTGTGCGGCCCTATGTGGGCATTCATCTGCGCATTTGGCTCT 660  
Qy 661 GACTGGAGAGCGCTGTGCTGCTGAGAGCGGACTGCGAGCTGCGACTTATGSGCC 720  
Db 661 GACTGGAGAGCGCTGTGCTGCTGAGAGCGGACTGCGAGCTGCGACTTATGSGCC 720  
Qy 721 TCTCGCAGTGTGTGGGCTACAGCGCGCAGCAGCGGCCCTCCACGATGACTATGTGC 780  
Db 721 TCTCGCAGTGTGTGGGCTACAGCGCGCAGCAGCGGCCCTCCACGATGACTATGTGC 780  
Qy 781 CTGCTGACCTGAAAGGAGATCAGAGGCTGTGAAGCTCTGGGTGAGTCTGATGCTGATGCC 840  
Db 781 CTGCTGACCTGAAAGGAGATCAGAGGCTGTGAAGCTCTGGGTGAGTCTGATGCTGATGCC 840  
Qy 841 CAGTGGCTTACGTTGCTGCTGATTCGAGAGTTATGTGCTTGGCTGAGCTTCAACAGCTCTTC 900  
Db 841 CAGTGGCTTACGTTGCTGCTGATTCGAGAGTTATGTGCTTGGCTGAGCTTCAACAGCTCTTC 900  
Qy 901 AAAGGGAAGTGAAGTGGTGAAGCTGAGGCTGAGGTGGCCAGGCTGACCTGTACATC 960  
Db 901 AAAGGGAAGTGAAGTGGTGAAGCTGAGGCTGAGGTGGCCAGGCTGACCTGTACATC 960  
Qy 961 CTCGGCCAAAGCCGACCTTTATTGGCAACTGTGTCTCTCTTCACTGCTTTTGTGAAG 1020  
Db 961 CTCGGCCAAAGCCGACCTTTATTGGCAACTGTGTCTCTCTTCACTGCTTTTGTGAAG 1020  
Qy 1021 CGGAGCGGAGCTTCCAGGAGAGGCGCTTCTTTTCTTGGCATGGACAGGCCCTTAAAG 1080  
Db 1021 CGGAGCGGAGCTTCCAGGAGAGGCGCTTCTTTTCTTGGCATGGACAGGCCCTTAAAG 1080

Qy 1081 CTGCGGGACGAGTCTTCTGATT 1100  
Db 1081 CTGCGGGACGAGTCTTCTGATT 1100  
RESULT 3  
US-09-774-954-4  
; Sequence 4, Application US/09774954  
; Publication No. US20040241645A1  
; GENERAL INFORMATION:  
; APPLICANT: Yang Wang, Michael W. Spellman  
; TITLE OF INVENTION: O-Fucosyltransferase  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genentech, Inc.  
; STREET: 1 DNA Way  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WinPatIn (Genentech)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/774,954  
; FILING DATE: 30-Jan-2001  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/978,741  
; FILING DATE: 26-NOV-1997  
; APPLICATION NUMBER: 08/792,498  
; FILING DATE: 31-JAN-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Svoboda, Craig G.  
; REGISTRATION NUMBER: 39,044  
; REFERENCE/DOCKET NUMBER: P1041P1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650/225-1489  
; TELEFAX: 650/952-9881  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1300 base pairs  
; TYPE: Nucleic Acid  
; STRANDEDNESS: Single  
; TOPOLOGY: Linear  
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:  
US-09-774-954-4

Query Match 99.9%; Score 1098.4; DB 11; Length 1300;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1099; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 ATGCCCCGGGGCTCTCTGGACCCGGCCGGTTACCTGCTCTACTGTCCTCCCTGATGGGGCGC 60  
Db 136 ATGCCCCGGGGCTCTCTGGACCCGGCCGGTTACCTGCTCTACTGTCCTCCCTGATGGGGCGC 195  
Qy 61 TTTGGGAACCAAGCCGATCACTTCTTGGGCTCTCTGGCAATTTGCAAGCTGCTAAACCGT 120  
Db 196 TTTGGGAACCAAGCCGATCACTTCTTGGGCTCTCTGGCAATTTGCAAGCTGCTAAACCGT 255  
Qy 121 ACCTTGGGCTGCTCCCTCTCTGGATTCAGTACAGCATCAAGCCCTCTTCCACCAACCTC 180  
Db 256 ACCTTGGGCTGCTCCCTCTCTGGATTCAGTACAGCATCAAGCCCTCTTCCACCAACCTC 315  
Qy 181 CATGTGCTCTACCAAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGCTCATC 240  
Db 316 CATGTGCTCTACCAAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGCTCATC 375  
Qy 241 AGCTTGGAGGATTTATGAGAGAGCTGGAGCCACCCACTGCGCCCTTGAGAAGCGGGTG 300  
Db 376 AGCTTGGAGGATTTATGAGAGAGCTGGAGCCACCCACTGCGCCCTTGAGAAGCGGGTG 435

301 GCATACCTGCTTTGAGTGGACCCAGCGAAGCCAGATGAAGACGTCGCCCATGAAG 360  
Db GCATACCTGCTTTGAGTGGACCCAGCGAAGCCAGATGAAGACGTCGCCCATGAAG 495  
Qy 361 GAAAGAAACCCCTTTGGCCCATCTTGGGATCAGTTCCTCATGTGAGTTTCAACAAGTCGGAG 420  
Db 496 GAAAGAAACCCCTTTGGCCCATCTTGGGATCAGTTCCTCATGTGAGTTTCAACAAGTCGGAG 555  
Qy 421 CTTTTTACAGCATTTCTTCTAGTCTCTTCTACAGAAACATGAAGCCAGAGATTTTCT 480  
Db 556 CTTTTTACAGCATTTCTTCTAGTCTCTTCTACAGAAACATGAAGCCAGAGATTTTCT 615  
Qy 481 CCAAGAGCAATCGGTGCTTCTGCTCCAGAGCCAGCCAGCCAGTCCCGTCTAGAA 540  
Db 616 CCAAGAGCAATCGGTGCTTCTGCTCCAGAGCCAGCCAGCCAGTCCCGTCTAGAG 675  
Qy 541 GAAACAGGCACTACAGAAATGATGTATGTGTCAGACGAAATGTGAAGACGGAGAG 600  
Db 676 GAAACAGGCACTACAGAAATGATGTATGTGTCAGACGAAATGTGAAGACGGAGAG 735  
Qy 601 GCCCAGATTATGCCACCTTTGTCGGCCCTATGTGGCAATTCATCTGGCATTTGGCTCT 660  
Db 736 GCCCAGATTATGCCACCTTTGTCGGCCCTATGTGGCAATTCATCTGGCATTTGGCTCT 795  
Qy 661 GACTGGAAGAACGCTGTGTCATGTCTGAAGACGGGACTGCAAGCTTCGCACTTCATGGCC 720  
Db 796 GACTGGAAGAACGCTGTGTCATGTCTGAAGACGGGACTGCAAGCTTCGCACTTCATGGCC 855  
Qy 721 TCTCGCAGTGTGTGGCTACAGCCGACAGCACAGCGCCGCCCTCACGATGACTATGTC 780  
Db 856 TCTCGCAGTGTGTGGCTACAGCCGACAGCACAGCGCCGCCCTCACGATGACTATGTC 915  
Qy 781 CTGCTGACCTGAGGAGATCAGAGGCTCTGAAGCTCTGGTGAGGTCCGTGAGTGC 840  
Db 916 CTGCTGACCTGAGGAGATCAGAGGCTCTGAAGCTCTGGTGAGGTCCGTGAGTGC 975  
Qy 841 CAGTCGCTTACGTTGCTACTGATTCGAGAGTTATGTGCTGAGCTCCAGACGCTTTC 900  
Db 976 CAGTCGCTTACGTTGCTACTGATTCGAGAGTTATGTGCTGAGCTCCAGACGCTTTC 1035  
Qy 901 AAAGGAAAGTGAAGTGTGAGGCTGAAAGCTGAGTGGCCAGGTGACCTGTPACATC 960  
Db 1036 AAAGGAAAGTGAAGTGTGAGGCTGAAAGCTGAGTGGCCAGGTGACCTGTPACATC 1095  
Qy 961 CTGGCCCAAGCCGACCACTTTATTTGGCAACTGTGTCTCTTCACTGCTTTGTGAAG 1020  
Db 1096 CTGGCCCAAGCCGACCACTTTATTTGGCAACTGTGTCTCTTCACTGCTTTGTGAAG 1155  
Qy 1021 CGGGAGCGGACCTCCAGGGAGGCGCTTCTTTCTTGGCAATGAGTGGCCAGGTGACCTG 1080  
Db 1156 CGGGAGCGGACCTCCAGGGAGGCGCTTCTTTCTTGGCAATGAGTGGCCAGGTGACCTG 1215  
Qy 1081 CTGGGGACGAGTTCTGATT 1100  
Db 1216 CTGGGGACGAGTTCTGATT 1235

## RESULT 4

US-09-774-954-5  
; Sequence 5, Application US/09774954  
; Publication No. US20040241645A1  
; GENERAL INFORMATION:  
; APPLICANT: Yang Wang, Michael W. Spellman  
; TITLE OF INVENTION: O-Fucosyltransferase  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genentech, Inc.  
; STREET: 1 DNA Way  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080

COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WinPatIn (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/774,954  
FILING DATE: 30-Jan-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/978,741  
FILING DATE: 26-Nov-1997  
APPLICATION NUMBER: 08/792,498  
FILING DATE: 31-Jan-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Svoboda, Craig G.  
REGISTRATION NUMBER: 39,044  
REFERENCE/DOCKET NUMBER: P1041P1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650/225-1489  
TELEFAX: 650/952-9881  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 11284 base pairs  
TYPE: Nucleic Acid  
STRANDEDNESS: Single  
TOPOLOGY: Linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 5:

US-09-774-954-5  
Query Match 99.9%; Score 1098.4; DB 11; Length 11284;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1099; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATGCCCGGGCTCTGGGACCGCGCGGTTACTGCTTACTGCCCCCTGCATGGGGCGC 60  
Db 4236 ATGCCCGGGCTCTGGGACCGCGCGGTTACTGCTTACTGCCCCCTGCATGGGGCGC 4295  
Qy 61 TTGGGAACACAGCCGATCACTCTTGGGCTCTGGCATTCGAAGCTGTAACCGT 120  
Db 4296 TTGGGAACACAGCCGATCACTCTTGGGCTCTGGCATTCGAAGCTGTAACCGT 4355  
Qy 121 ACCTTGGTGTCCTCTCTTGGATTTGAGTACAGAGTCAAGGCTCTTTTACCAACCTC 180  
Db 4356 ACCTTGGTGTCCTCTCTTGGATTTGAGTACAGAGTCAAGGCTCTTTTACCAACCTC 4415  
Qy 181 CATGTGTCCTACCAAGAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGTCATC 240  
Db 4416 CATGTGTCCTACCAAGAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGTCATC 4475  
Qy 241 AGCTTGGAGGATTTTCATGGAGAGCTGGCAACCACCTGGCCCTCGAGAAGCGGGTG 300  
Db 4476 AGCTTGGAGGATTTTCATGGAGAGCTGGCAACCACCTGGCCCTCGAGAAGCGGGTG 4535  
Qy 301 GCATACCTGTTTGGAGTGGCAGCCAGAGGAGCCAGATGAAGACGTCGCCCATGAAG 360  
Db 4536 GCATACCTGTTTGGAGTGGCAGCCAGAGGAGCCAGATGAAGACGTCGCCCATGAAG 4595  
Qy 361 GAAGAAACCCCTTTGGCCCATCTTGGATCAGTTCATGTGAGTTTCAACAAGTCGGAG 420  
Db 4596 GAAGAAACCCCTTTGGCCCATCTTGGATCAGTTCATGTGAGTTTCAACAAGTCGGAG 4655  
Qy 421 CTTTTTACAGCATTTCTTCTAGTCTCTTCTACAGAAACATGAAGCCAGAGATTTTCT 480  
Db 4656 CTTTTTACAGCATTTCTTCTAGTCTCTTCTACAGAAACATGAAGCCAGAGATTTTCT 4715  
Qy 481 CCAAGAGCAATCGGTGCTTGGCTCCAGAGCCAGCCAGCCAGTCCCGTCTTAGAA 540  
Db 4716 CCAAGAGCAATCGGTGCTTGGCTCCAGAGCCAGCCAGCCAGTCCCGTCTTAGAG 4775  
Qy 541 GAAACAGGCACTACAGAAATGATGTGTCAGACGAAATGTGAAGACGGAGAG 600  
Db 4776 GAAACAGGCACTACAGAAATGATGTGTCAGACGAAATGTGAAGACGGAGAG 4835









Publication No. US20050118625A1  
; GENERAL INFORMATION:  
; APPLICANT: Wyeth  
; APPLICANT: Mounts, William  
; TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION ASSOCIATED WITH  
; TITLE OF INVENTION: HUMAN OSTEOARTHRITIS AND HUMAN PROTEASES  
; FILE REFERENCE: 031896-043000 (AM 101081)  
; CURRENT APPLICATION NUMBER: US/10/956.157  
; CURRENT FILING DATE: 2004-10-04  
; NUMBER OF SEQ ID NOS: 319805  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 4751  
; LENGTH: 1506  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-956-157-4751

Query Match 42.9%; Score 471.4; DB 22; Length 1506;  
Best Local Similarity 99.8%; Pred. No. 1.9e-137;  
Matches 472; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 ATGCCCGCGGCTCTCTGGGACCGCGGCTTACCTGCTCTACTGCCCCCTGATGGGCGC 60  
Db 119 ATGCTGCGGGCTCTCTGGGACCGCGGCTTACCTGCTCTACTGCCCCCTGATGGGCGC 178  
Qy 61 TTGGGACAGCCGATCACTTCTGGGCTCTCTGGGCTCTGCAAGCTGCTAAACCGT 120  
Db 179 TTGGGACAGCCGATCACTTCTGGGCTCTCTGGGCTCTGCAAGCTGCTAAACCGT 238  
Qy 121 ACCTTGGCTGCTCTCTCTGGGATTTAGATACAGGATCACAAAGCTCTCTTACCAACCTC 180  
Db 239 ACCTTGGCTGCTCTCTCTGGGATTTAGATACAGGATCACAAAGCTCTCTTACCAACCTC 298  
Qy 181 CATGTGCTCTACAGAAAGTACTTCAAGCTGGAGCCCTCCAGGCTTTACCATCGGTCATC 240  
Db 299 CATGTGCTCTACAGAAAGTACTTCAAGCTGGAGCCCTCCAGGCTTTACCATCGGTCATC 358  
Qy 241 ACCTTGGAGGATTTCTATGAGAGAGCTGGCACCACCTGGCCCCCTGAGAAGCGGTG 300  
Db 359 ACCTTGGAGGATTTCTATGAGAGAGCTGGCACCACCTGGCCCCCTGAGAAGCGGTG 418  
Qy 301 GCATCTGCTTTGAGTGGGCGCCAGCCAGCCAGATTAAGAGCTGCCCATGAG 360  
Db 419 GCATCTGCTTTGAGTGGGCGCCAGCCAGCCAGATTAAGAGAGCTGCCCATGAG 478  
Qy 361 GAAGAAACCCCTTTGGCCCATTTCTGGGATCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420  
Db 479 GAAGAAACCCCTTTGGCCCATTTCTGGGATCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 538  
Qy 421 CTTTTTACAGGCATTTCTTTCAGTGTCTTCTACAGAGAAACAATGGAGCCAGAG 473  
Db 539 CTTTTTACAGGCATTTCTTTCAGTGTCTTCTACAGAGAAACAATGGAGCCAGAG 591

RESULT 10  
US-764-891-6032  
; Sequence 6032, Application US/09764891  
; Publication No. US20030077808A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC006  
; CURRENT APPLICATION NUMBER: US/09/764,891  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 10231  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 6032  
; LENGTH: 10331  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-764-891-6032

Query Match 22.8%; Score 251.2; DB 10; Length 10331;  
Best Local Similarity 97.0%; Pred. No. 9.4e-68;  
Matches 256; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
Qy 656 GCTCTGACTGGAAGACGGCTGTGCCATCTGCTGAAGGACGGGACTCGAGCTCGCACTTCA 715  
Db 2480 GCTCTCCCTAGAGAACGGCTGTGCCATCTGCTGAAGGACGGGACTCGAGCTCGCACTTCA 2539  
Qy 716 TGGCCTCTCCGCACTGTGTGGGCTACAGCCGAGCAGAGGGGCCCTCAGCATGACTA 775  
Db 2540 TGGCCTCTCCGCACTGTGTGGGCTACAGCCGAGCAGAGGGGCCCTCAGCATGACTA 2599  
Qy 776 TGTGCTCTGCTGACCTCAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGTGCCTGG 835  
Db 2600 TGTGCTCTGCTGACCTCAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGTGCCTGG 2659  
Qy 836 ATGCCAGTGTGCTTACGTTTCTACTGATTCGAGAGATTATGCTCGTGAAGCTCCCAACAGC 895  
Db 2660 ATGCCAGTGTGCTTACGTTTCTACTGATTCGAGAGATTATGCTCGTGAAGCTCCCAACAGC 2719  
Qy 896 TCTTCAAAAGGGAAGTGAAGGTGG 919  
Db 2720 TCTTCAAAAGGGAAGTGAAGGTGG 2743

RESULT 11  
US-09-918-995-19437  
; Sequence 19437, Application US/09918995  
; Publication No. US20030073623A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; FILE REFERENCE: 20411-756  
; CURRENT APPLICATION NUMBER: US/09/918,995  
; CURRENT FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US/09/235,076  
; PRIOR FILING DATE: 1999-01-20  
; NUMBER OF SEQ ID NOS: 38054  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 19437  
; LENGTH: 479  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-918-995-19437

Query Match 21.2%; Score 233; DB 10; Length 479;  
Best Local Similarity 100.0%; Pred. No. 1.7e-62;  
Matches 233; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 868 GAGAGTTATGTGCTGAGCTCCCAACAGCTCTTCAAAGGGAAGGTGAAGTGTGAGCCCTG 927  
Db 31 GAGAGTTATGTGCTGAGCTCCCAACAGCTCTTCAAAGGGAAGGTGAAGTGTGAGCCCTG 90  
Qy 928 AAGCTGAGGTGGCCGAGGTCGACCTGTACATCTCTCGCCAAAGCCGACCTTTATTGGC 987  
Db 91 AAGCTGAGGTGGCCGAGGTCGACCTGTACATCTCTCGCCAAAGCCGACCTTTATTGGC 150  
Qy 988 AACTGTGTCTCTCTTCACTGCTTTTGAAGCGGAGCGGACCTCCAGGGGAGGCGG 1047  
Db 151 AACTGTGTCTCTCTTCACTGCTTTTGAAGCGGAGCGGACCTCCAGGGGAGGCGG 210  
Qy 1048 TCTTCTTTCTTCGGCATGAGAGGCCCTTAAAGCTCGGGGACGAGTTCTGATT 1100  
Db 211 TCTTCTTTCTTCGGCATGAGAGGCCCTTAAAGCTCGGGGACGAGTTCTGATT 263

RESULT 12  
US-09-918-995-22744  
; Sequence 22744, Application US/09918995  
; Publication No. US20030073623A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.

```
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: FROM VARIOUS cDNA LIBRARIES
; CURRENT APPLICATION NUMBER: US/09/918,995
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 22744
; LENGTH: 477
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc_feature
; LOCATION: (1)...(477)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-22744
```

```
Query Match      20.2%; Score 221.8; DB 10; Length 477;
Best Local Similarity 96.3%; Pred. No. 5.9e-59;
Matches 237; Conservative 0; Mismatches 8; Indels 1; Gaps 1;

QY 855 TCCTACTGATCCGAGAGTTATGTGCTGAGCTCCACAGCTCTTCAAAGGGAAGTGAA 914
Db 11 TGATCTGAGCNCATAGAGATATGTGCTGAGCTCC-ACAGTTCTTCAAAGGGAAGTGAA 69

QY 915 GGTGTGAGCCTGAAGCCTGAGGTGGCCAGGTGCACTGTACATCTCTCGGCGAAGCCGA 974
Db 70 GGTGTGAGCCTGAAGCCTGAGGTGGCCAGGTGCACTGTACATCTCTCGGCGAAGCCGA 129

QY 975 CCACCTTATGGCAACTGTGTCTCTCTTCACTGCTCTTGTGAAGCGGAGCGGACCT 1034
Db 130 CCACCTTATGGCAACTGTGTCTCTCTTCACTGCTCTTGTGAAGCGGAGCGGACCT 189

QY 1035 CCAGGGAGGCGCTTCTTCTTCTGCGCATGAGAGGCGCCCTTAAGCTGCGGAGCGATT 1094
Db 190 CCAGGGAGGCGCTTCTTCTTCTGCGCATGAGAGGCGCCCTTAAGCTGCGGAGCGATT 249

QY 1095 CTGATT 1100
Db 250 CTGATT 255
```

```
RESULT 13
US-11-097-143-18161
; Sequence 18161, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
```

```
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18161
; LENGTH: 1209
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-11-097-143-18161
```

```
Query Match      15.1%; Score 166.4; DB 26; Length 1209;
Best Local Similarity 51.5%; Pred. No. 2.4e-41;
Matches 563; Conservative 0; Mismatches 486; Indels 45; Gaps 6;

QY 17 GGGACCCGGCGGTTTACCTGCTCTACTGCCCTGTCATGGGGCGCTTTGGGAACCAAGCGCG 76
Db 80 GCGATCCCAATGGCTTACCTACTCTGCTGCTGATGGAGCGCTTTGGCAACCAAGCGCG 139

QY 77 ATCACTTCTTGGGCTCTCTGGGCAATTTGCAAGCTGTCTAAACGTTACCTTGGCTGTCCCTC 136
Db 140 ACCACTTCTTGGGATCATTTGGCCTTTCGCAAGGCGCTTTAATCGCACCTGTATCTCCGCG 199

QY 137 CTTGATTTGAGTACCAGCATCACAGCCTCTCTTTCCACCAACTCCATGTCCTCAACGA 196
Db 200 CGTGGTGGAGT-----ATCGTGGGTGAATCGCGATCCCGTCAGGTACCGTTCAACA 253

QY 197 AGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGCTCATCAGCTTGGAGGATTTCA 256
Db 254 CATATTTTGAAGTGGAGCCCTGAAGGAATACCATCGCTCATCACCATGGCAGATTTCA 313

QY 257 TGGAGAGCTGGCAACCCACCTGGCCCCCTGAGAACGGGTGGCATACTGCTTTTGAGG 316
Db 314 TGTGGCACCTGGCGGACGACATTTTGGCCAGAATCGGAGCGAGTGTCTATTTTGTACTACAAG 373

QY 317 -----TGGCAGCCCGAGGAGCCAGATAAGAGACGTGCCCATGAAGG 361
Db 374 AACGATATAGCCTTCAGCAGGAGAGAACGATTCAGACAAGCCCAATTGCACGCGCAAGG 433

QY 362 AAGGAAACCCCTTTGGGCCATTTCTGGGATCAGTTTCTGATGTTTCAACAGTCGGAGC 421
Db 434 ATGCAATCTTTTGTGCTCTTTTGGGACATTTTCAATAGACTTTTGTGCGGTGAGAT 493

QY 422 TTTTACAGGCAATTTCTTCAAGTGTCTCTACAG-----AGAAACAATGGAGCCAGA 472
Db 494 TCTATGGCGCACTTCAATTTTGTATGTCATCATAGCAACGAGGCTGCCAAGTGGCAGACA 553

QY 473 GATTTTCTCAAAGGAACATCCGCTGCTTGGCTTGGCCCTGCGCAGGAGCCCGACCGCTCCCG 532
Db 554 AATATCTCTGCAAGATCATATCCCGTACTCGCGTTCACCGGAGCTCCGCTAGTTTCTCTG 613

QY 533 TCCTAGAAGAACACAGGCGCACTACAGAAAGTATGTTGTCAGACCGAAATGGTGAAGA 592
Db 614 TTCAGCTAGAGAACTGCAAGCTGCGCGCTACTTTCAGTGGAGTCAACGGTATAGGGAAG 673

QY 593 CGGAGAGGCCAGATTTATGTCGCCACCTTGTTCGCGC---CCTATGTGGGCAATTCATCTGC 649
Db 674 CATCTAAGGATTTTCATCCGAGAGCAGTTTGCCTCGGGGTGCTTTTGGGCACTCATCTGC 733

QY 650 GCATTTGCTCTGACTGGAGAAACCGCTGTCCTGCAATGCTGAAGACGG-----GACTG 700
Db 734 GCAACGGTATCGATTGGGTGAGAGCCCTGTGAGACGCTCAAGGATAGCCAGCATCTGTTTG 793

QY 701 CAGGCTCGCACTTTCATGGCCTCTCCGCGCTGTGTGGGCTACAGCGCGAGCAGCGGCCC 760
Db 794 CCTCGCGCAGTGTCTGGGCTATAAAATGNAACGTGGTGCACTCTACCGGAGCTCTGCA 853

QY 761 CCCTCAGATGACTATATGCTGCTGCTGACCTGGAAGGAGATCCAGAGGCTGTGAAGCTCT 820
Db 854 TGCCCTCAAAGGAGGCGCATCATCCGCCAGCTTAAAGAGAACCATTTAAGAACCTGCGCCAAA 913

QY 821 GGGTGAGTCTGCTGATGCCCGCAGTCTGCTACGTTGCTACTGATTCGAGAGGTTATGTGC 880
Db 914 CTCAGCCGGCAACGAAATCAAAATCAGTTTTCGTGGCGTCCAGACTCCAAATCACATGATTG 973

QY 881 CTGAGCTCCAACAGCTCTTCAAAGGGAAGGTGAAGGTGGTGAGCTGGAAGCTGAGGTGG 940
```

Db 974 GTGAACAAACGCGCCCTTAGTCGCATGGGCATAGTGTGCACAAAGCTGCGGAGGATG 1033  
Qy 941 ---CCCAAGTGCACCTGTATACCTCTCGCCAAAGCCGACCACTTTATTTGGCAACTGTGTCT 997  
Db 1034 ATCCTTACCTGGACTTTGGCCATCTCGACAGTCCGACCACTTTATTCGGCAACTGTATAT 1093  
Qy 998 CTTCTCTTCACTGCTTTGTGAAGCGGAGCGGACCTCCAGGGAGGCGCTCTTCTTCT 1057  
Db 1094 CTTCTTACTCGGCATTCGAAAAAGGAACGAGATGTGCAGCGTTTTCATCGTACTTCT 1153  
Qy 1058 TCGGCATGGACAGG 1071  
Db 1154 GGGGATCCCCAAG 1167

RESULT 14  
US-11-097-143-18160/c  
; Sequence 18160, Application US/11097143  
; Publication No. US20050208558A1  
; GENERAL INFORMATION:  
; APPLICANT: Venter, J. Craig  
; APPLICANT: et al.  
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID  
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE  
; TITLE OF INVENTION: DROSOPHILA GENES.  
; FILE REFERENCE: CL000728  
; CURRENT APPLICATION NUMBER: US/11/097,143  
; CURRENT FILING DATE: 2005-04-04  
; PRIOR FILING DATE: 1999-10-05  
; PRIOR FILING DATE: 1999-10-19  
; PRIOR FILING DATE: 1999-10-19  
; PRIOR FILING DATE: 1999-10-19  
; PRIOR FILING DATE: 1999-10-28  
; PRIOR FILING DATE: 1999-11-12  
; PRIOR FILING DATE: 1999-11-12  
; PRIOR FILING DATE: 1999-12-28  
; PRIOR FILING DATE: 1999-12-28  
; PRIOR FILING DATE: 2000-01-12  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR FILING DATE: 2000-03-23  
; NUMBER OF SEQ ID NOS: 43008  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 18160  
; LENGTH: 3264  
; TYPE: DNA  
; ORGANISM: DROSOPHILA  
US-11-097-143-18160

Query Match 13.7%; Score 150.4; DB 26; Length 3264;  
Best Local Similarity 50.9%; Pred. No. 3.7e-36;  
Matches 541; Conservative 0; Mismatches 476; Indels 45; Gaps 6;  
Qy 49 TGCATGGGCGCTTTGGACACGAGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAG 108  
Db 2098 TTCAAGAGCGCTTTGGACACGAGCCGACCACTTCTGGGATCATTTGGCTTCGCCAAG 2039  
Qy 109 CTGCTAACCCTGACTTGGCTGTCCTCTTGGATTTAGTACAGCATCAAGCCTTCT 168  
Db 2038 GCGCTTAATCGCACCTTGATCTCCGCGCGTGGGGAGT-----ATCtTAGGGGTGAA 1985  
Qy 169 TTCACCAACCTTCATGTGTCTTACACAGAGTACTTCAAGCTGGAGCCCTCCAGGCTTAC 228  
Db 1984 CTGGATCCCGTCAAGTACCGTTCACACATATTTGAAGTGGAGCCCTCGAAGAAATAC 1925  
Qy 229 CATCGGCTCATCAGCTTGGAGGATTTTCATGGAGAGCTGGCACCCACCCACCTGGGCCCT 288  
Db 1924 CATCGGCTCATCAGCTTGGAGGATTTTCATGTGGCACCTGGCCGACGACATTTTGGCCAGAA 1865

Qy 289 GAGAACGGGTGGGCATATCTGCTTTGAGG-----TGCGACCCCGCAGCAAGC 333  
Db 1864 TCGGAGCGAGTGTATCTTTGCTACAAGGAACGATATAGCCTTCAGCAGGAGAAGCAAT 1805  
Qy 334 CCAGATAGAGAGAGCTGCCCCCATGAAGAGGAAGAACCCCTTTGGCCCATTTCTGGGATCAG 393  
Db 1804 CCAGACAGGCCAATTCGCCACGCCAAGGATGCGAATCTTTTGGTCCCTTTTGGGACACT 1745  
Qy 394 TTTCAATGTAGTGTTCACAAAGTCCGAGAGCTTTTATACAGGCATTTCTTTCAGTGTCTTCCTAC 453  
Db 1744 TTTCAATAGACTTTTGGGGTTCAGAGTTCTATGCCCACTTCATTTTGTATGTGCATCAT 1685  
Qy 454 AG-----AGAACAAATGGAGCCAGAGATTTTCTCCAAAGGAACATCCGGTGTCTGCC 504  
Db 1684 AGCAACGAGGCTGCCAAGTGGCAGACCAATATCTTCAGATCATATCCCTACTCGCG 1625  
Qy 505 CTGCCAGAGCCCGCAGCCAGTTCGCCCTCTAGAGAAACACAGGSCCATACAGAAGTAC 564  
Db 1624 TTCACCGAGCTCCGGCTAGTTTCTCTGTTAGCTAGAGAACTGCAAGCTGCGCGCTAC 1585  
Qy 565 ATGATATGCTCAGACGAAATGTGAAGACGGGAGAGCCCGAGATTATGCCACCTTGTTC 624  
Db 1564 TTGCAGTGGAGTCAACGGTATAGGGAAGCATCTTAAGGATTTTCATCCGAGAGCAGTTGCTT 1505  
Qy 625 CGG---CCCTATGTGGGCATTTTCATCTGCGCATTTGGCTCTGACTGGAAGAACCCCTGTGCC 681  
Db 1504 CGGGGTGCTTTTGGGCGATTCATCTGCGCAACGGTATCGATTGGGTGGAGAGCCCTGTGAG 1445  
Qy 682 ATGCTGAAGAGACGG-----GACTGCAGGCTCGCACTTCATGGCCCTCTCCGAGTGT 732  
Db 1444 CACGTCAAGGATAGCCAGCATCTGTTTGCCTCGCGCAGTGTCTGGGCTATATAAATGAA 1385  
Qy 733 GTGGGCTACAGCCGCGACGACAGCGGCCCTCTCAGCATGACTATGTGCTGTGCTGACCTG 792  
Db 1384 CGTGTGCACCTCTACCCGGAGCTCTGCATGCCCTCCAAAGGAGCGCATCATCCGCCAGCTA 1325  
Qy 793 AAGGAGATCCAGAGGCGCTGGAAGCTCTGGGTGAGTGTGCTGAGTCCCGAGTCCAGTCTGCTAC 852  
Db 1324 AAGAGAACCATTAAGAACCTGCGCAAACTCAGCCGCGACCAACGAATCAATCAGTTTC 1265  
Qy 853 GTTCTACTGATTCGAGAGTATGTGCTGAGTCTCAACAGCTCTTCAAGGGAGGTG 912  
Db 1264 GTGGGTGAGTCTCAATCATCATGATTTGGTGAATTAACACGCGCTTGTGCTGCTGGC 1205  
Qy 913 AAGGTGTGAGCTTGAAGCTGAGGTGG---CCAGGTGCACTGTGATCATCTCTCGGCCAA 969  
Db 1204 ATCAGTGTGCACAAAGTCCGAGGAGTATCTTACCTGGACTTGGCCATTTCTGGACAG 1145  
Qy 970 GCCGACCACTTTATTTGGCAACTGTGTCTCTCTCTTCACTGCTTGTGAAGCGGAGCGG 1029  
Db 1144 TCGAACCACTTTATTCGGCAACTGTATATCTCTTACTCGGCATTCGAAAAAAGGGAACGA 1085  
Qy 1030 GACCTCCAGGCGAGCGCTCTTCTTTCTTTCGGCATGGACAGG 1071  
Db 1084 GATGTGCACGGTTCATTCGATCTTTCGGGATTCGCCAAG 1043

RESULT 15  
US-09-783-590-11501  
; Sequence 11501, Application US/09783590  
; Patent No. US20020110850A1  
; GENERAL INFORMATION:  
; APPLICANT: Dillon, Patrick J.  
; APPLICANT: Haseltine, William A.  
; APPLICANT: Li, Haodong  
; APPLICANT: Rosen, Craig A.  
; APPLICANT: Ruben, Steven M.  
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2  
; FILE REFERENCE: PO-16.2C1  
; CURRENT APPLICATION NUMBER: US/09/783,590  
; CURRENT FILING DATE: 2000-02-15  
; PRIOR APPLICATION NUMBER: 08/420,856  
; PRIOR FILING DATE: 1995-04-12

```

; PRIOR APPLICATION NUMBER: 08/346,731
; PRIOR FILING DATE: 1994-11-21
; NUMBER OF SEQ ID NOS: 12485
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11501
; LENGTH: 402
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (9)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (19)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (22)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (46)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (272)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (301)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (320)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (333)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (342)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (374)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (386)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (390)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (400)
; OTHER INFORMATION: n equals a,t,g, or c
; US-09-783-590-11501

```

```

Query Match      10.4%; Score 114.6; DB 9; Length 402;
Best Local Similarity 83.9%; Pred. No. 3.4e-25;
Matches 177; Conservative 0; Mismatches 23; Indels 11; Gaps 4;

Qy 178 CTCATGTGCTCTACCAAGTACTTCAAGTGGAGCCCTCCAGGCTTACCATCGGGTC 237
Db 161 CTCATGTGCTCTACCAAGTACTTCAAGTGGAGCCCTCCAGGCTTACCATCGGGTC 220

Qy 238 ATCAGCTT-GGAGGATTTTCATGGAGAGCTGGCACCACCCAGCTGGCCCTT---GAGAA 293
Db 221 ATCAGCTTGGGAGGATTTTCATGGAGAGCTGGCACCACCCAGCTGGCCCTTGGAGAGC 280

Qy 294 GCGGGTGGCATACTGCTTTGA--GGTGGCAGCCAGCAAGCCAGAT-----AAGAAGA 346
Db 281 GGGGTGGGCATATGCTTTTNNAGTGGGCGCCAGCCGCGGCGCCAGATTAAAGNAGGACG 340

Qy 347 CGTGCCCATGAAGAGCAACCCCTTTGG 377
Db 341 TNGACCCCATGAAGAGCAACCCCTTTG 371

```

Search completed: November 9, 2005, 15:51:17  
Job time : 988.698 secs

THIS PAGE BLANK (USPTO)



Qy 61 TTTGGGAACACGCGCATCTCTTGGGCTCTCTGGCATTTGCAAGCTGCTAAACCGT 120  
Db 61 TTTGGGAACACGCGCATCTCTTGGGCTCTCTGGCATTTGCAAGCTGCTAAACCGT 120  
Qy 121 ACCTTGGCTGTCCCTCTCTTGGATTCAGTACACAGCATCAACAGCTCTCTTTACCAACCTC 180  
Db 121 ACCTTGGCTGTCCCTCTCTTGGATTCAGTACACAGCATCAACAGCTCTCTTTACCAACCTC 180  
Qy 181 CATGTGCTTACCAAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGCTCATC 240  
Db 181 CATGTGCTTACCAAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGCTCATC 240  
Qy 241 AGCTTGGAGATTTTCAGAGAGCTGCGACCGACCATCTGGGCCCTCGAGACGGGTG 300  
Db 241 AGCTTGGAGATTTTCAGAGAGCTGCGACCGACCATCTGGGCCCTCGAGACGGGTG 300  
Qy 301 GCATCTGCTTTGAGTGGAGCGCCAGCGAAGCCGACATGAAGAACGCTGCCCCATGAAG 360  
Db 301 GCATCTGCTTTGAGTGGAGCGCCAGCGAAGCCGACATGAAGAACGCTGCCCCATGAAG 360  
Qy 361 GAAGGAACCCCTTTGGCCCATTTCTGGGATCAGTTTTCATGTGAGTTTCAACAAGTCCGAG 420  
Db 361 GAAGGAACCCCTTTGGCCCATTTCTGGGATCAGTTTTCATGTGAGTTTCAACAAGTCCGAG 420  
Qy 421 CTTTTTACAGGCATTTCTTCACTGCTTCTTACAGAGAACATGAGAGCCAGAGATTTTCT 480  
Db 421 CTTTTTACAGGCATTTCTTCACTGCTTCTTACAGAGAACATGAGAGCCAGAGATTTTCT 480  
Qy 481 CCAAGGAACATCCGGTGTCTTCCCTCCAGGAGCCCGCCAGCTTCCCGTCTAGAA 540  
Db 481 CCAAGGAACATCCGGTGTCTTCCCTCCAGGAGCCCGCCAGCTTCCCGTCTAGAA 540  
Qy 541 GAAACACAGGCCATCACAGAGTACATGTATGTTCAGACGAATGTGAAGCCGGAG 600  
Db 541 GAAACACAGGCCATCACAGAGTACATGTATGTTCAGACGAATGTGAAGCCGGAG 600  
Qy 601 GCCAGATTCATGCCCATCTGTGCGGCCCTATGTGGGATTCATCTGGGCTTGGCTCT 660  
Db 601 GCCAGATTCATGCCCATCTGTGCGGCCCTATGTGGGATTCATCTGGGCTTGGCTCT 660  
Qy 661 GACTGGAAGACCGCTGTGCGGCCCTATGTGGGATTCATCTGGGCTTGGCTCT 720  
Db 661 GACTGGAAGACCGCTGTGCGGCCCTATGTGGGATTCATCTGGGCTTGGCTCT 720  
Qy 721 TCTCCGAGTGTGTGGGCTACAGCGGAGCTGTGAAGCTCTGGGTGAGGTGCGTGGATGCC 780  
Db 721 TCTCCGAGTGTGTGGGCTACAGCGGAGCTGTGAAGCTCTGGGTGAGGTGCGTGGATGCC 780  
Qy 781 CTGCTGACCTGAAGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGGTGCGTGGATGCC 840  
Db 781 CTGCTGACCTGAAGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGGTGCGTGGATGCC 840  
Qy 841 CAGTCGCTACGTTGCTACTGATTCAGAGAGTTATGTGCTGAGCTCCACAGCTCTTC 900  
Db 841 CAGTCGCTACGTTGCTACTGATTCAGAGAGTTATGTGCTGAGCTCCACAGCTCTTC 900  
Qy 901 AAAGGGAAGTGAAGTGTGAGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 960  
Db 901 AAAGGGAAGTGAAGTGTGAGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 960  
Qy 961 CTGGCCAAAGCCGACCATTTTATGGGAACTGTGTCTCTCTCTCTCTCTCTCTCTCTCTCT 1020  
Db 961 CTGGCCAAAGCCGACCATTTTATGGGAACTGTGTCTCTCTCTCTCTCTCTCTCTCTCTCT 1020  
Qy 1021 CGGAGGCGGACCTTCAGGGGAGGCGCTCTTCTTCTTGGGATGACAGGCGCCCTTAAG 1080  
Db 1021 CGGAGGCGGACCTTCAGGGGAGGCGCTCTTCTTCTTGGGATGACAGGCGCCCTTAAG 1080  
Qy 1081 CTGGGGACGAGTTCTGATT 1100  
Db 1081 CTGGGGACGAGTTCTGATT 1100

RESULT 2  
US-09-333-729A-16  
; Sequence 16, Application US/09333729A  
; Patent No. 6270987  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Yang  
; APPLICANT: Spellman, Michael W.  
; TITLE OF INVENTION: O-Fucosyltransferase  
; FILE REFERENCE: P1041PDI-Substitute  
; CURRENT APPLICATION NUMBER: US/09/333, 729A  
; PRIOR FILING DATE: 1999-06-15  
; PRIOR APPLICATION NUMBER: US 08/798, 741  
; PRIOR FILING DATE: 1997-11-26  
; NUMBER OF SEQ ID NOS: 21  
; SEQ ID NO 16  
; LENGTH: 1100  
; TYPE: DNA  
; ORGANISM: Homo Sapien  
US-09-333-729A-16  
  
Query Match 100.0%; Score 1100; DB 3; Length 1100;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 ATGCCCGGGCTCTCTGGAGCCCGGCTTACCTGTCTACTGCCCTGCAATGGGGCG 60  
Db 1 ATGCCCGGGCTCTCTGGAGCCCGGCTTACCTGTCTACTGCCCTGCAATGGGGCG 60  
Qy 61 TTTGGGAACACGCGCATCACTTCTTGGGCTCTCTGGCATTTGCAAGCTGCTAAACCGT 120  
Db 61 TTTGGGAACACGCGCGATCACTTCTTGGGCTCTCTGGCATTTGCAAGCTGCTAAACCGT 120  
Qy 121 ACCTTGGCTGTCCCTCTCTTGGATTCAGTACACAGCATCAAGGCTCTCTTACCAACCTC 180  
Db 121 ACCTTGGCTGTCCCTCTCTTGGATTCAGTACACAGCATCAAGGCTCTCTTACCAACCTC 180  
Qy 181 CATGTGCTTACCAAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGCTCATC 240  
Db 181 CATGTGCTTACCAAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGCTCATC 240  
Qy 241 AGCTTGGAGATTTTCAGAGAGCTGCGACCGACCATCTGGGCCCTCGAGACGGGTG 300  
Db 241 AGCTTGGAGATTTTCAGAGAGCTGCGACCGACCATCTGGGCCCTCGAGACGGGTG 300  
Qy 301 GCATCTGCTTTGAGTGGAGCGCCAGCGAAGCCGACATGAAGAACGCTGCCCCATGAAG 360  
Db 301 GCATCTGCTTTGAGTGGAGCGCCAGCGAAGCCGACATGAAGAACGCTGCCCCATGAAG 360  
Qy 361 GAAGGAACCCCTTTGGCCCATTTCTGGGATCAGTTTTCATGTGAGTTTCAACAAGTCCGAG 420  
Db 361 GAAGGAACCCCTTTGGCCCATTTCTGGGATCAGTTTTCATGTGAGTTTCAACAAGTCCGAG 420  
Qy 421 CTTTTTACAGGCATTTCTTCACTGCTTCTTACAGAGAACATGAGAGCCAGAGATTTTCT 480  
Db 421 CTTTTTACAGGCATTTCTTCACTGCTTCTTACAGAGAACATGAGAGCCAGAGATTTTCT 480  
Qy 481 CCAAGGAACATCCGGTGTCTTCCCTCCAGGAGCCCGCCAGCTTCCCGTCTAGAA 540  
Db 481 CCAAGGAACATCCGGTGTCTTCCCTCCAGGAGCCCGCCAGCTTCCCGTCTAGAA 540  
Qy 541 GAAACACAGGCCATCACAGAGTACATGTATGTTCAGACGAATGTGAAGCCGGAG 600  
Db 541 GAAACACAGGCCATCACAGAGTACATGTATGTTCAGACGAATGTGAAGCCGGAG 600  
Qy 601 GCCAGATTCATGCCCATCTGTGCGGCCCTATGTGGGATTCATCTGGGCTTGGCTCT 660  
Db 601 GCCAGATTCATGCCCATCTGTGCGGCCCTATGTGGGATTCATCTGGGCTTGGCTCT 660  
Qy 661 GACTGGAAGACCGCTGTGCGGCCCTATGTGGGATTCATCTGGGCTTGGCTCT 720  
Db 661 GACTGGAAGACCGCTGTGCGGCCCTATGTGGGATTCATCTGGGCTTGGCTCT 720  
Qy 721 TCTCCGAGTGTGTGGGCTACAGCGGAGCTGTGAAGCTCTGGGTGAGGTGCGTGGATGCC 780





QY 1081 CTGCGGACGAGTTCTGATT 1100  
Db 1081 CTGCGGACGAGTTCTGATT 1100

## RESULT 4

US-09-333-729A-2  
; Sequence 2, Application US/09333729A  
; Patent No. 6270987  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Yang  
; APPLICANT: Spellman, Michael W.  
; TITLE OF INVENTION: O-Fucosyltransferase  
; FILE REFERENCE: P104IPDI-Substitute  
; CURRENT APPLICATION NUMBER: US/09/333,729A  
; CURRENT FILING DATE: 1999-06-15  
; PRIOR APPLICATION NUMBER: US 08/798,741  
; PRIOR FILING DATE: 1997-11-26  
; NUMBER OF SEQ ID NOS: 21  
; SEQ ID NO 2  
; LENGTH: 1514  
; TYPE: DNA  
; ORGANISM: Homo Sapien  
US-09-333-729A-2

Query Watch 100.0%; Score 1100; DB 3; Length 1514;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATGCCCGGGGCTCTGGGACCGCGCGGTACCTGCTCTACTGCCCCCTGCGATGCGGCGC 60  
Db 1 ATGCCCGGGGCTCTGGGACCGCGCGGTACCTGCTCTACTGCCCCCTGCGATGCGGCGC 60  
QY 61 TTGGGAACACGAGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAAGCTGTAACCGGT 120  
Db 61 TTGGGAACACGAGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAAGCTGTAACCGGT 120  
QY 121 ACCTTGGCTGTCCTCTCTGGATTCAGTACAGCATCACAAAGCTCTCTTACCAACCTC 180  
Db 121 ACCTTGGCTGTCCTCTCTGGATTCAGTACAGCATCACAAAGCTCTCTTACCAACCTC 180  
QY 181 CATGTGCTTACCAAGACTTCAAGTGGAGCCCTCCAGGCTTACCATCGGTCATC 240  
Db 181 CATGTGCTTACCAAGACTTCAAGTGGAGCCCTCCAGGCTTACCATCGGTCATC 240  
QY 241 AGCTTGGAGGATTTATGAGAGAGTGGCACCCACCTGGCCCCCTGAGAAGCGGCTG 300  
Db 241 AGCTTGGAGGATTTATGAGAGAGTGGCACCCACCTGGCCCCCTGAGAAGCGGCTG 300  
QY 301 GCATATCTCTTGAAGTGGCAGCCAGCAAGCCAGATTAAGAAAGCTGCCCCATGAAG 360  
Db 301 GCATATCTCTTGAAGTGGCAGCCAGCAAGCCAGATTAAGAAAGCTGCCCCATGAAG 360  
QY 361 GAAGGAACCCCTTGGCCCATCTGGGATCAGTTTCATGTGAGTTTCAACAGTCGGAG 420  
Db 361 GAAGGAACCCCTTGGCCCATCTGGGATCAGTTTTCATGTGAGTTTCAACAGTCGGAG 420  
QY 421 CTTTATACAGGATTTCTTCTAGTGTCTTCTACAGAAACAATGGAGCCAGAGATTTCT 480  
Db 421 CTTTATACAGGATTTCTTCTAGTGTCTTCTACAGAAACAATGGAGCCAGAGATTTCT 480  
QY 481 CCAAGGAACATCCGCTGCTTGGCCCTGCGAGGAGCCCGAGCTTCCCGCTCTCTAGAA 540  
Db 481 CCAAGGAACATCCGCTGCTTGGCCCTGCGAGGAGCCCGAGCTTCCCGCTCTCTAGAA 540  
QY 541 GAACACAGGCCACTACAGAGTACATGTTGTTGTCAGAGAAATGTTGAAGACGGAGAG 600  
Db 541 GAACACAGGCCACTACAGAGTACATGTTGTTGTCAGAGAAATGTTGAAGACGGAGAG 600  
QY 601 GCCAGATTCATGCCACTGTCGGGCTTATGTTGGGATTCATCTGGGATTTGCTCT 660  
Db 601 GCCAGATTCATGCCACTGTCGGGCTTATGTTGGGATTCATCTGGGATTTGCTCT 660

QY 661 GACTGGGAAGAACGCTGTGCCATGTGAAGGACGGGACTGAGGCTCGCACTTCATGSCC 720  
Db 661 GACTGGGAAGAACGCTGTGCCATGTGAAGGACGGGACTGAGGCTCGCACTTCATGSCC 720  
QY 721 TCTCCGAGTGTGTGGGCTACAGCCGACAGCAGCGGCCCTCTACGATGACTATGTGC 780  
Db 721 TCTCCGAGTGTGTGGGCTACAGCCGACAGCAGCGGCCCTCTACGATGACTATGTGC 780  
QY 781 CTGCTGACCTGAAGGAGATCCAGAGGCTCTGGAAGCTCTGGGTGAGTCTGGTATGCC 840  
Db 781 CTGCTGACCTGAAGGAGATCCAGAGGCTCTGGAAGCTCTGGGTGAGTCTGGTATGCC 840  
QY 841 CAGTCGCTTACGTTGCTACTGATTCGAGAGTTATGTCCTGAGCTCCAAAGCTCTTC 900  
Db 841 CAGTCGCTTACGTTGCTACTGATTCGAGAGTTATGTCCTGAGCTCCAAAGCTCTTC 900  
QY 901 AAAGGAAAGGTGAAGGTGGTGAAGCTTGAAGCTGAGGTGGGCCAGGTGCACTGTACATC 960  
Db 901 AAAGGAAAGGTGAAGGTGGTGAAGCTTGAAGCTGAGGTGGGCCAGGTGCACTGTACATC 960  
QY 961 CTGCGCCAAAGCCGACCACTTTATTGGCAACTGTGTCTCTCTTCACTGCTTTGTGAAG 1020  
Db 961 CTGCGCCAAAGCCGACCACTTTATTGGCAACTGTGTCTCTCTTCACTGCTTTGTGAAG 1020  
QY 1021 CGGAGCGGAGCTTCCAGGCGGAGCGCTCTTCTTTCGCGATGGACAGGCCCTTAAG 1080  
Db 1021 CGGAGCGGAGCTTCCAGGCGGAGCGCTCTTCTTTCGCGATGGACAGGCCCTTAAG 1080  
QY 1081 CTGCGGAGCAGGTTCTGATT 1100  
Db 1081 CTGCGGAGCAGGTTCTGATT 1100

## RESULT 5

US-08-978-741-4  
; Sequence 4, Application US/08978741  
; Patent No. 6100076  
; GENERAL INFORMATION:  
; APPLICANT: Yang Wang, Michael W. Spellman  
; TITLE OF INVENTION: O-Fucosyltransferase  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSER: Genentech, Inc.  
; STREET: 1 DNA Way  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WinPatIn (Genentech)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/978,741  
; FILING DATE: 26-Nov. 6100076-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/792498  
; FILING DATE: 31  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Svoboda, Craig G.  
; REGISTRATION NUMBER: 39,044  
; REFERENCE/DOCKET NUMBER: P1041P1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650/225-1489  
; TELEFAX: 650/952-9881  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1300 base pairs  
; TYPE: Nucleic Acid  
; STRANDEDNESS: Single  
; TOPOLOGY: Linear

US-08-978-741-4

```
Query Match          99.9%; Score 1098.4; DB 3; Length 1300;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1099; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGCCCGGGGCTCTGGGACCCGCGGTACCTGCTCTACTGCCCCCTGCATGGGGCGC 60
DB 136 ATGCCCGGGGCTCTGGGACCCGCGGTACCTGCTCTACTGCCCCCTGCATGGGGCGC 195

QY 61 TTGCGGAACCGGCGGATCACTCTTGGGCTCTGCGCATTTGCAAGCTGCTAAACCGT 120
DB 196 TTGCGGAACCGGCGGATCACTCTTGGGCTCTGCGCATTTGCAAGCTGCTAAACCGT 255

QY 121 ACCTTGGCTGTCCTCTTGGATTGAGTACAGCATCAAGACCTCTCTTACCAACCTC 180
DB 256 ACCTTGGCTGTCCTCTTGGATTGAGTACAGCATCAAGACCTCTCTTACCAACCTC 315

QY 181 CATGTGCTCTACCAAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATGGGTCTATC 240
DB 316 CATGTGCTCTACCAAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATGGGTCTATC 375

QY 241 AGCTTGGAGGATTTCAAGAGAGTGGCAACCCACCTGCCCCCTGAGAGCGGGTG 300
DB 376 AGCTTGGAGGATTTCAAGAGAGTGGCAACCCACCTGCCCCCTGAGAGCGGGTG 435

QY 301 GCATACCTGCTTGGGTGGCGCCAGCGAGCCAGATAGAGACCTGCCCATGAAG 360
DB 436 GCATACCTGCTTGGGTGGCGCCAGCGAGCCAGATAGAGACCTGCCCATGAAG 495

QY 361 GAAGAAACCCCTTTGGCCCATTTGGGATCAGTTTCAATGTAGTTTCAACAAGTCGGAG 420
DB 496 GAAGAAACCCCTTTGGCCCATTTGGGATCAGTTTCAATGTAGTTTCAACAAGTCGGAG 555

QY 421 CTTTITTAAGGATTTCTTCAAGTCTCTTACAGAGAACTATGAGACGACGATTTTCT 480
DB 556 CTTTITTAAGGATTTCTTCAAGTCTCTTACAGAGAACTATGAGACGACGATTTTCT 615

QY 481 CCAAGGAACATCCGGTCTGCGCCCTGCGAGGCCAGCCAGTCTCCCGTCTAGAA 540
DB 616 CCAAGGAACATCCGGTCTGCGCCCTGCGAGGCCAGCCAGTCTCCCGTCTAGAA 675

QY 541 GAACACAGGCCACTACAGAGTACATGTATGGTCAGACGAATGGTGAAGACGGGAG 600
DB 676 GAACACAGGCCACTACAGAGTACATGTATGGTCAGACGAATGGTGAAGACGGGAG 735

QY 601 GCCCAGATTATGCCCACTTGTGCGGCCCTATGTGGGCAATCATCTGCGCATTTGGCTCT 660
DB 736 GCCCAGATTATGCCCACTTGTGCGGCCCTATGTGGGCAATCATCTGCGCATTTGGCTCT 795

QY 661 GACTGGAAGACCGCTGTGCCATGCTGAAGACGGGACTGCGAGGCTCGCACTTCATGGCC 720
DB 796 GACTGGAAGACCGCTGTGCCATGCTGAAGACGGGACTGCGAGGCTCGCACTTCATGGCC 855

QY 721 TCTCGCAGTGTGTGGGTACAGCGCAGCAGCGGCCCTCCACGATGACTATGTGC 780
DB 856 TCTCGCAGTGTGTGGGTACAGCGCAGCAGCGGCCCTCCACGATGACTATGTGC 915

QY 781 CTGCTGACCTGAAGGAGATCCAGAGGCTGTGAAGCTCTGGGTAGGTCTGCTGATGCC 840
DB 916 CTGCTGACCTGAAGGAGATCCAGAGGCTGTGAAGCTCTGGGTAGGTCTGCTGATGCC 975

QY 841 CAGTCGGTCTAGTTGCTACTGATTTCCAGAGTATATGCTGAGCTCCACAGCTCTTC 900
DB 976 CAGTCGGTCTAGTTGCTACTGATTTCCAGAGTATATGCTGAGCTCCACAGCTCTTC 1035

QY 901 AAAGGAAGGTGAAGGTGGTGAAGCTCAGAGTGGCCCGCCAGGTCGACTGTACATC 960
DB 1036 AAAGGAAGGTGAAGGTGGTGAAGCTCAGAGTGGCCCGCCAGGTCGACTGTACATC 1095

QY 961 CTCGCCCAAGCGGACCACTTTATTGGCAACTGTGTCTCTCTTCACTGCTTTGTGAAG 1020
DB 1096 CTCGCCCAAGCGGACCACTTTATTGGCAACTGTGTCTCTCTTCACTGCTTTGTGAAG 1155
```

```
QY 1021 CGGAGCGGACCTCCAGGGAGGCGGCTCTTCTTTTCGTCATGACAGGCCCTCTAAG 1080
DB 1156 CGGAGCGGACCTCCAGGGAGGCGGCTCTTCTTTTCGTCATGACAGGCCCTCTAAG 1215

QY 1081 CTGCGGACGAGTTCTGATT 1100
DB 1216 CTGCGGACGAGTTCTGATT 1235

RESULT 6
US-09-333-729A-6
; Sequence 6, Application US/09333729A
; Patent No. 6270987
; GENERAL INFORMATION:
; APPLICANT: Wang, Yang
; APPLICANT: Spellman, Michael W.
; TITLE OF INVENTION: O-Fucosyltransferase
; FILE REFERENCE: F1041P1D1-Substitute
; CURRENT APPLICATION NUMBER: US/09/333,729A
; CURRENT FILING DATE: 1999-06-15
; PRIOR APPLICATION NUMBER: US 08/798,741
; PRIOR FILING DATE: 1997-11-26
; NUMBER OF SEQ ID NOS: 21
; SEQ ID NO 6
; LENGTH: 1300
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Plasmid insert.
US-09-333-729A-6

Query Match          99.9%; Score 1098.4; DB 3; Length 1300;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1099; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGCCCGGGGCTCTGGGACCCGCGGTACCTGCTCTACTGCCCCCTGCATGGGGCGC 60
DB 136 ATGCCCGGGGCTCTGGGACCCGCGGTACCTGCTCTACTGCCCCCTGCATGGGGCGC 195

QY 61 TTGCGGAACCGGCGGATCACTCTTGGGCTCTCTGCGCATTTGCAAGCTGCTAAACCGT 120
DB 196 TTGCGGAACCGGCGGATCACTCTTGGGCTCTCTGCGCATTTGCAAGCTGCTAAACCGT 255

QY 121 ACCTTGGCTGTCCTCTTGGATTGAGTACAGCATCAAGACCTCTCTTACCAACCTC 180
DB 256 ACCTTGGCTGTCCTCTTGGATTGAGTACAGCATCAAGACCTCTCTTACCAACCTC 315

QY 181 CATGTGCTCTACCAAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATGGGTCTATC 240
DB 316 CATGTGCTCTACCAAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATGGGTCTATC 375

QY 241 AGCTTGGAGGATTTCAAGAGAGTGGCAACCCACCTGCCCCCTGAGAGCGGGTG 300
DB 376 AGCTTGGAGGATTTCAAGAGAGTGGCAACCCACCTGCCCCCTGAGAGCGGGTG 435

QY 301 GCATACCTGCTTGGGTGGCGCCAGCGAGCCAGATAGAGACCTGCCCATGAAG 360
DB 436 GCATACCTGCTTGGGTGGCGCCAGCGAGCCAGATAGAGACCTGCCCATGAAG 495

QY 361 GAAGAAACCCCTTTGGCCCATTTGGGATCAGTTTCAATGTAGTTTCAACAAGTCGGAG 420
DB 496 GAAGAAACCCCTTTGGCCCATTTGGGATCAGTTTCAATGTAGTTTCAACAAGTCGGAG 555

QY 421 CTTTITTAAGGATTTCTTCAAGTCTCTTACAGAGAACTATGAGACGACGATTTTCT 480
DB 556 CTTTITTAAGGATTTCTTCAAGTCTCTTACAGAGAACTATGAGACGACGATTTTCT 615

QY 481 CCAAGGAACATCCGGTCTGCGCCCTGCGAGGCCAGCCAGTCTCCCGTCTAGAA 540
DB 616 CCAAGGAACATCCGGTCTGCGCCCTGCGAGGCCAGCCAGTCTCCCGTCTAGAA 675

QY 541 GAACACAGGCCACTACAGAGTACATGTATGGTCAGACGAATGGTGAAGACGGGAG 600
```







Db 607 TTCACTAGAGAACTCAAGCTGCAAGCGCTACTTTCAGTGGAGTCAACGGTATAGGGAAG 666  
Qy 593 CGGAGAGGCCAGATTATGTCCTCCACCTTGTCCGCG---CTATGTGGCATTATCTGCG 649  
Db 667 CATCTAAGGATTTTATCCGAGAGCAGTGTGCTCGGGGTGCCCTTTTGGGCAATTCATCTGC 726  
Qy 650 GCATTGGCTCTGACTGGAAGAACGCTGTGCCATGCTGAAGACGG-----GACTG 700  
Db 727 GCACGGTATCGATTGGGTGAGAGCTGTGAGCAGCTCAAGGATAGCAGCATCTGTGTTG 786  
Qy 701 CAGGCTCGCACTTATGCGCTCTCCGCAAGTGTGCGGTACAGCGCCAGCAGCGGCC 760  
Db 787 CCTCGCCGAGTGTGCGCTATAAAATGAACGTGGTGCACTTACCGGAGCTCTGCA 846  
Qy 761 CCCTCAGATGACTATGTCCTGCTGCTCCGCAAGTGTGCGGTACAGAGGCTGTGAAGCTCT 820  
Db 847 TGCCCTCCAAAGGAGCGCATCATCCGCCAGCTAAAGAGAACCATTAAGAACGTCGCCCAA 906  
Qy 821 GGGTGAGTTCGCTGGATGCCAGTCCGCTGCTACGTTGTCTACTGATTCCGAGAGTATGTGC 880  
Db 907 CTCAGCCGACNACGAATCAATCAGTTTTCGTGCGTCAAGCTCCATCACATGATG 966  
Qy 881 CTGAGCTCCAAAGCTCTTCAAAGGAAG---TGAAGGTGGTGAGCCTGAAGCCTGAGG 937  
Db 967 GTGAACCTAAACACGCGCTTACTGCGCATGGGCATCAGTGTGCACAAGCTGACGGAGGATG 1026  
Qy 938 TGGCCAGGTGCACTGTATCATCTCCGCAAGCGCACCTTATTGCGCACTGTGCT 997  
Db 1027 ATCTTACTGACTTGGCCATTTCTCGGACAGTTCGAACCACTTTATCGGCACTGTATAT 1086  
Qy 998 CTTCTTCACTGCTCTTGTGAAGCGGAGCGGACCTCCAGGGAGGCGCTCTCTTCT 1057  
Db 1087 CTTCTTACTGCGCATCTGTAAAGGAGACGAGATGTGACGGTTTTCATCGTACTTCT 1146  
Qy 1058 TCGGCATGACAGG 1071  
Db 1147 GGGGATTCGCCAAG 1160

## RESULT 11

US-09-270-767-30507

; Sequence 30507, Application US/09270767

; Patent No. 6703491

; GENERAL INFORMATION:

; APPLICANT: Homburger et al.

; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*

; FILE REFERENCE: File Reference: 7326-094

; CURRENT APPLICATION NUMBER: US/09/270,767

; CURRENT FILING DATE: 1999-03-17

; NUMBER OF SEQ ID NOS: 62517

; SOFTWARE: Patent in Ver. 2.0

; SEQ ID NO 30507

; LENGTH: 722

; TYPE: DNA

; ORGANISM: *Drosophila melanogaster*

US-09-270-767-30507

Query Match 5.7%; Score 62.2; DB 4; Length 722;  
Best Local Similarity 49.5%; Pred. No. 2.2e-08;  
Matches 225; Conservative 0; Mismatches 218; Indels 12; Gaps 2;  
Qy 629 CCTATGTGGGCAATTCATCTGGCATTGCTGCTGGAAGAACGCCCTGTGCCATGTGA 688  
Db 108 CTTTGTGGGCAATTCATCTGGCATTGCTGCTGGAAGAACGCCCTGTGCCATGTGA 167  
Qy 689 AGGACGG-----GACTGCGAGCTCGCACTTCATGCGCTCTCCGAGTGTGGGCT 739  
Db 168 AGGATAGCAGCATCTGTTGCTGCGCAGTGTCTGGGCTATAAATGAACGTGTG 227  
Qy 740 ACAGCCGAGCAGCAGCGGCCCTTCAGATGACTATGTGCTGCTGACCTGAAGGAGA 799  
Db 228 CACTCTACCGGAGCTCTGCATGCGCTCCAAAGGAGCGCATCATCGCCAGCTAAAGAA 287

Qy 800 TCCAGAGGCTGTGAAGCTCTGGGTGAGGTGCGTGGATGCCAGTCCGTCTACGTTGCTA 859  
Db 288 CCATTAAAGACGTGGCCCAAACTCAGCCGAGACAGAAATCAATCAGTTTCTGTGGGCT 347  
Qy 860 CTGATTCGAGAGATTATGTGCTGAGCTCCCAACAGCTCTTCAAAGGGAAG---TGAAGG 916  
Db 348 CAGACTCCAATCACATGATTGGTGAACCTAAACAGGCCCTTAGTCGCATGGSCATCAGTG 407  
Qy 917 TGGTGAGCCTGAAGCCTGAGGTGGCCAGGTGCGACCTGTATCATCTCGGCCAAGCGGACC 976  
Db 408 TGCAACAGCTCAGCGAGGATGATCCTTACCTGGACTTGGCCATTCTCGACAGTCAAGC 467  
Qy 977 ACTTTATTGGCAACTGTGTCTCTCTTCTTCTTTCGCGCATGACAGG 1071  
Db 468 ACTTTATCGGCAACTGTATATCTTCTTACTCGCATTCGTAAGGGAACAGATGTGC 527  
Qy 1037 AGGGAGGCCGCTTCTTCTTTCGCGCATGACAGG 1071  
Db 528 ACGGTTTTCCATCGTACTTCTGGGATTCCCAAG 562

## RESULT 12

US-09-270-767-28905/c

; Sequence 28905, Application US/09270767

; Patent No. 6703491

; GENERAL INFORMATION:

; APPLICANT: Homburger et al.

; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*

; FILE REFERENCE: File Reference: 7326-094

; CURRENT APPLICATION NUMBER: US/09/270,767

; CURRENT FILING DATE: 1999-03-17

; NUMBER OF SEQ ID NOS: 62517

; SOFTWARE: Patent in Ver. 2.0

; SEQ ID NO 28905

; LENGTH: 998

; TYPE: DNA

; ORGANISM: *Drosophila melanogaster*

US-09-270-767-28905

Query Match 5.3%; Score 58.6; DB 4; Length 998;  
Best Local Similarity 60.2%; Pred. No. 3.2e-07;  
Matches 97; Conservative 0; Mismatches 64; Indels 0; Gaps 0;  
Qy 911 TGAAGTGTGCTGAGCTGAGCCTGAAGCCTGAGGTGGCCAGGTGACATCTCGGCCAAG 970  
Db 896 TCAGTGTGCACAAGCTGACGGAGATGATCTTACCTGGACTTGGCCATTCTCGACAGT 837  
Qy 971 CCGACCACTTTTATGGCAACTGTGTCTCTCTTCACTGCTTTGTGAAGCGGAGCGGG 1030  
Db 836 CGAACCACTTTTATCGGCAACTGTATATCTTCTTACTCGGCATTCTGTAAGGGAACGAG 777  
Qy 1031 ACCTCCAGGGGAGCGCTCTTCTTCTTTCGCGCATGACAGG 1071  
Db 776 ATGTGCACGGTTTTCATCGTACTTCTGGGATTCCCAAG 736

## RESULT 13

US-09-270-767-13022/c

; Sequence 13022, Application US/09270767

; Patent No. 6703491

; GENERAL INFORMATION:

; APPLICANT: Homburger et al.

; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*

; FILE REFERENCE: File Reference: 7326-094

; CURRENT APPLICATION NUMBER: US/09/270,767

; CURRENT FILING DATE: 1999-03-17

; NUMBER OF SEQ ID NOS: 62517

; SOFTWARE: Patent in Ver. 2.0

; SEQ ID NO 13022

; LENGTH: 2034

; TYPE: DNA

; ORGANISM: *Drosophila melanogaster*

US-09-270-767-13022

```
Query Match          5.3%; Score 58.6; DB 4; Length 2034;
Best Local Similarity 60.2%; Pred. No. 4.8e-07;
Matches 97; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

Qy 911 TGAAGTGGTGCAGCTGTAAGCCTGAGGTGGCCAGGTGACATCTCGGCCAAG 970
Db 1932 TCAGTGTGCAAGCTGACGGAGGATGATCCTTACCTGGACTTGGCCATCTCGGACAGT 1873

Qy 971 CGGACCACTTATGCAACTGTCTCTCTCTCACTGCTTGTGAAGCGGAGCGGG 1030
Db 1872 CGAACCACTTATCGGCACTGTATATCTCTTACTCGGCATTCGTAAGGAACGAG 1813

Qy 1031 ACCTCCAGGGAGGCGCTCTTCTTCTTCGCGCATGGAACAG 1071
Db 1812 ATGTCACGGSTTTTCCATGCTACTTCTGGGGATTCCCCCAAG 1772
```

RESULT 14  
US-09-949-016-12127/c  
Sequence 12127, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
FILE REFERENCE: CL001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
CURRENT FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 12127  
LENGTH: 19227  
TYPE: DNA  
ORGANISM: Human  
US-09-949-016-12127

```
Query Match          5.1%; Score 56; DB 4; Length 19227;
Best Local Similarity 100.0%; Pred. No. 1.1e-05;
Matches 56; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGCCCGCGGGCTCTGGGACCGCGGTTACCTGCTTACTGCCCCCTGCATGGG 56
Db 1718 ATGCCCGCGGGCTCTGGGACCGCGGTTACCTGCTTACTGCCCCCTGCATGGG 1663
```

RESULT 15  
US-09-949-016-16285/c  
Sequence 16285, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
FILE REFERENCE: CL001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
CURRENT FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 16285

```
LENGTH: 19228
TYPE: DNA
ORGANISM: Human
US-09-949-016-16285

Query Match          5.1%; Score 56; DB 4; Length 19228;
Best Local Similarity 100.0%; Pred. No. 1.1e-05;
Matches 56; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGCCCGCGGGCTCTGGGACCGCGGTTACCTGCTTACTGCCCCCTGCATGGG 56
Db 1718 ATGCCCGCGGGCTCTGGGACCGCGGTTACCTGCTTACTGCCCCCTGCATGGG 1663

Search completed: November 9, 2005, 10:46:08
Job time : 202.044 secs
```